THE SHADOW VALUE OF ATTENTION: RECOMMENDATION RANKING WITH INVENTORY DYNAMICS

PANTELIS LOUPOS AND CAN URGUN

Abstract. Recommendation engines and inventory systems are typically optimized separately: the former allocate scarce consumer attention to stimulate demand, while the latter manage the operating costs created by that demand. We study recommendation order as a dynamic exposure-allocation problem for an online retailer selling N substitutable products. Demand depends on product demand potential and recommendation position; the firm earns margins, incurs holding and backlog costs, and replenishes inventory at fixed-plus-linear cost. We formulate the problem as a discounted continuous-time Markov decision process whose state space grows exponentially in N. To obtain a tractable policy, we relax the coupled ranking problem into a multi-mode restless-bandit problem with one-copy-per-position constraints. We show that replenishment has a threshold structure, the relaxed single-product problem admits a partition over inventory states, and the relaxation yields a closed-form index for each product-position-inventory state. The index measures the shadow value of assigning recommendation attention by internalizing demand potential, margin, inventory state, backlog cost, replenishment cost, and the incremental attention created by a higher position. The resulting policy converts an exponentially large dynamic program into a scalable assignment rule. In a four-product benchmark solvable by exact dynamic programming, a bootstrapped index closely tracks the optimum, with an average optimality gap of 3%, while reducing weighted-average cost by 92% relative to demand-only ranking and by 94% relative to myopic control. In large-catalog simulations, the index improves on static demand- and margin-based rankings, reducing discounted cost per item by up to 65.6% under stressed demand.

Date: June 2026.
Key words and phrases. Recommendation systems; inventory management; algorithmic attention; e-commerce operations; restless multi-mode bandits .
Loupos: Graduate School of Business, University of California, Davis, ploupos@ucdavis.edu. Urgun: Department of Economics, The University of North Carolina at Chapel Hill, curgun@unc.edu.

1. Introduction

Recommendation systems allocate a scarce economic resource: consumer attention. In online retail, the order in which products are displayed does not merely reveal what consumers are likely to want; it changes which products receive incremental demand from the recommendation system. A product placed near the top of a list receives more exposure and generates more purchase opportunities. A product placed lower in the list may still have intrinsic appeal, but receives less recommendation-induced exposure. Recommendation order is therefore an allocation decision, not simply a prediction problem or a user-interface choice.

This allocation problem is dynamic. Promoting a product today changes not only current revenue but also the future inventory state. If the product has ample stock, additional exposure may be valuable because it converts inventory into margin. If the product is near stockout or already backlogged, the same exposure may be costly because it increases shortage penalties, accelerates replenishment pressure, and reduces the firm's ability to serve future demand. Thus, a ranking rule that is optimal for current demand or current revenue need not be optimal for the firm once the continuation value of inventory is taken into account.

A simple example illustrates the core trade-off. Consider two highly substitutable products, one with ample stock and one near depletion. If the products have similar margins and demand potential, allocating the higher recommendation position to the well-stocked product is natural: the firm can convert attention into sales without creating large future scarcity costs. But this logic does not imply that scarce products should always be demoted. A scarce product may still merit prominent placement if its margin or demand potential is sufficiently high. The relevant comparison is therefore not inventory alone, margin alone, or predicted demand alone. It is the inventory-adjusted marginal value of additional exposure: the expected margin generated by the incremental demand from a higher position, net of the future cost of depleting inventory. A rule based only on predicted demand, click-through probability, or one-period expected revenue can fail because it prices the immediate benefit of exposure but not the continuation-value cost imposed by the induced sale. Recommendation ranking is therefore a dynamic demand-allocation problem across constrained assets, not simply a prediction or revenue-scoring problem.

The same logic extends beyond online retail. A restaurant server who recommends a dish should not only consider which item the customer is most likely to enjoy, but also whether the kitchen can produce it without creating congestion at the grill, overloading a chef, or delaying other orders. A hotel clerk recommending room upgrades, a sales representative steering customers across substitutable products, or a service platform directing demand across workers faces the same economic problem: a recommendation shifts demand toward an underlying constrained asset. The constrained asset may be physical inventory, but it may also be labor, production capacity, equipment, or service capacity. In these simpler settings, the firm may recommend only one item or steer demand toward one option. Online recommendation ranking is the high-dimensional version of the same problem: the platform does not merely choose whether to recommend a single item, but chooses an ordered allocation of attention across many products, each linked to a constrained operating state.

This economic tension also has an organizational counterpart. In many firms, the systems that allocate customer attention are governed separately from the systems that manage the constrained assets affected by that attention. In online retail, recommendation teams are often evaluated on predictive or short-run commercial metrics, such as relevance, click-through, conversion, or immediate expected revenue, while operations teams bear the consequences of the demand that those rankings create: depletion, backlog, restocking pressure, and fulfillment failures. Analogous misalignments arise whenever customer-facing employees or algorithms steer demand toward resources managed elsewhere in the organization, such as kitchen capacity, service labor, delivery slots, or production equipment. The result is an internal misalignment. The demand-shaping system controls exposure, but its local objective does not fully price the future cost imposed on constrained operating assets. Inventory-aware ranking can therefore be viewed as one instance of a broader governance problem for demand-shaping systems: the firm must align the objective used to allocate exposure with the shadow value of the constrained assets that exposure uses or depletes.

We study this problem in an online retail setting in which the firm repeatedly chooses recommendation rankings and product-specific replenishment actions for N substitutable products. The demand system has two primitives. First, products differ in item-level demand potential: some products are more likely than others to generate demand when they receive consumer attention. We denote this product-specific demand potential by θ_i. Second, consumers face costly attention when inspecting an ordered recommendation list. In Section 3.1, we derive this attention process from a costly-inspection model in which consumers choose how deeply to inspect the list. This yields position-specific attention weights, with higher positions receiving weakly more attention. Combining these two primitives gives a tractable demand representation in which a product's demand rate depends on both its own demand potential and the attention generated by its assigned position. This structure is consistent with ordered-search and position-effect models in online markets (e.g., Varian, 2007; Athey and Ellison, 2011).

The key economic object is the inventory-adjusted marginal value of demand. A sale of product i generates an immediate margin, m_i, but it also reduces the future inventory of that product by one unit. That inventory change has a continuation-value cost. When inventory is plentiful, this cost may be small. When inventory is scarce or backlogged, it may dominate the margin because additional demand raises expected shortage, backlog, replenishment, or service-reliability costs. A static ranking rule tends to allocate high-attention positions to products with high demand potential, high margin, or high one-period expected revenue. A dynamic rule instead allocates attention according to the value of exposure after accounting for the future inventory cost created by the induced demand. The same product can therefore be worth promoting in one inventory state and worth demoting in another.

We develop a theoretical framework by embedding ordered recommendation exposure in a continuous-time inventory-control model. The firm repeatedly chooses a recommendation ranking and product-specific replenishment actions. A ranking assigns products to positions and thereby determines their demand intensities. A product-specific replenishment action restores the selected product to its full stock level at fixed-plus-linear cost. Inventory evolves stochastically as demand arrives and as products are replenished. The firm earns margins on fulfilled demand and incurs holding and backlog costs. Its objective is to minimize expected discounted net cost, or equivalently to maximize discounted profit net of inventory and replenishment costs. The resulting control problem links ranking and replenishment: exposure affects current demand, current demand changes inventory, and inventory determines the future value of additional exposure.

The exact dynamic program is well defined but computationally infeasible for realistic catalogs. The joint inventory state space grows exponentially in the number of products, and the ranking decision couples products through the one-product-per-position constraint. Even in the four-product benchmark used in our experiments, with twelve inventory levels per product, the exact state space contains 20,736 states and 28 actions per state. Adding products quickly makes exact dynamic programming impractical. A useful theory of recommendation ranking therefore cannot stop at the Bellman equation. It must deliver a decision rule that preserves the relevant economic trade-offs while remaining implementable at catalog scale.

To obtain such a rule, we formulate the problem as a multi-mode restless-bandit problem and apply a Whittle-style Lagrangian relaxation (Whittle, 1988; Niño-Mora, 2001). In this formulation, each product is an arm, recommendation positions are modes that allocate different amounts of attention to the product, and replenishment is an additional action. Relaxing the global ranking constraint introduces a shadow price of attention and decomposes the coupled problem into single-product subproblems. Because this step relaxes the original constraint, the resulting policy is a heuristic for the original dynamic program rather than an exact solution. The relaxation nevertheless transforms the system-level ranking problem into product-level marginal-value calculations, which can then be recombined through an assignment rule.

Our main analytical result is a closed-form index for each product–position–inventory state. The index measures the shadow value of assigning a product to a particular ranking position at a given inventory level. It internalizes both the incremental attention generated by the position, and therefore the incremental demand, and the inventory-related costs that this induced demand creates. In a standard index policy, the decision rule is simple: allocate the resource to the arm with the highest index. In our multi-position ranking problem, the same logic applies with one additional step. Because products must be assigned to distinct positions, the firm computes the product–position indices and solves a static assignment problem that matches products to positions. This converts the original exponentially large dynamic program into a tractable static optimization over closed-form index values.

We evaluate the index policy in two settings. First, in a four-product benchmark where the exact dynamic program is tractable, we solve the discounted problem by policy iteration and compare the index policy with the exact optimum and with natural heuristics. A bootstrapped version of the index reduces weighted-average cost by 92% relative to a demand-only heuristic and by 94% relative to myopic control. It also closely tracks the exact optimum, attaining an average optimality gap of about 3% across the state space. The residual gap reflects the cross-product coordination that only the full joint-state dynamic program can exploit; it is modest at every inventory level and vanishes in scarcity and backlog states, where the ranking–replenishment trade-off is most economically consequential.

Second, we study large catalogs where exact dynamic programming is infeasible. In simulations aligned with the continuous-time Markov decision process (CT-MDP), the index policy improves on static demand- and margin-based rankings. In a finite-horizon ranking-only regime, it reduces discounted cost per item by 1.9% at $N = 1{,}000$, by 9.2% at $N = 200$, and by up to 65.6% under stressed demand. In an instantaneous-reset regime, the same qualitative pattern remains, with smaller gaps because replenishment absorbs part of the cost of poor exposure allocation. These results show that the index improves performance by reallocating exposure away from products whose current demand contribution is outweighed by the future inventory cost of additional demand, and toward products with higher inventory-adjusted exposure value.

The paper makes three contributions. First, it reframes recommendation order as a dynamic allocation of scarce attention under operational scarcity. The firm is not merely choosing which products are relevant to consumers; it is assigning a demand-generating shared resource across products with different margins, replenishment costs, backlog costs, and inventory shadow values. Second, it identifies the economically relevant ranking criterion: the inventory-adjusted marginal value of exposure. This criterion differs from product-specific demand potential, predicted demand, margin, and immediate expected revenue because it incorporates the continuation value of inventory. Third, it provides a tractable implementation of this logic. It is easy to verify that any optimal policy must have a threshold level below which replenishment is optimal (henceforth, called replenishment threshold), i.e., the reset region is downward closed. We then show that the relaxed single-product problem admits a “partition" structure over inventory states, where rankings are monotone. Building on the methodology in Nino-Mora (2006); Niño-Mora (2007), we show that the multi-mode restless-bandit relaxation yields a closed-form marginal product index (MPI)[1]. The resulting index policy converts the original high-dimensional dynamic program into a scalable assignment rule for ranking products in real time.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3.1 provides the attention microfoundation for the demand decomposition. Section 3 develops the dynamic model and formulates the exact continuous-time control problem. Section 4 introduces the restless-bandit relaxation, derives the

[1]Throughout the paper, we use “MPI" and “index" interchangeably to refer to the marginal product index.

closed-form index, and characterizes the induced ranking policy. Section 5 evaluates the policy in both the exact small-system benchmark and large-scale simulations aligned with the continuous-time Markov decision process. Section 6 concludes.

2. Literature Review

Our paper relates to four literatures: recommendation systems and algorithmic attention allocation, inventory-constrained demand management, strategic resource allocation in platforms and organizations, and restless-bandit index policies. These literatures share a common concern with decisions that allocate demand, attention, or capacity under system-level constraints. Recommendation models study how firms direct attention across products; inventory and revenue-management models study how firms manage scarce capacity across demand opportunities; platform and organizational models show how local decisions and local metrics can create costs elsewhere in the system; and restless-bandit models provide tractable priority rules for high-dimensional dynamic allocation problems. Our contribution is to connect these ideas in a model of inventory-aware recommendation ranking. We treat ordered recommendation exposure as a dynamic control that shifts demand across products, changes future inventory states, and interacts with product-specific replenishment. The resulting problem differs from standard recommendation models, which usually optimize relevance or short-run business value; from assortment and revenue-management models, which usually choose prices, capacities, or offered sets rather than ordered display positions; and from standard restless-bandit applications, which typically do not involve multi-position attention allocation with product-specific inventory and replenishment.

The first related literature studies recommendation systems. Traditional recommenders use collaborative filtering, content-based methods, and matrix-factorization approaches to infer consumer preferences from historical interactions (Linden et al., 2003; Adomavicius and Tuzhilin, 2005; Koren et al., 2009; Schafer et al., 2007). More recent models use richer sequential, contextual, and business-objective formulations to predict clicks, purchases, engagement, revenue, profit, or customer lifetime value (De Biasio et al., 2023, 2024). This literature provides the main toolkit for estimating which products consumers are likely to value. These models typically treat the recommendation score as a predictive object used to rank products at a point in time. Inventory enters, if at all, as an availability filter, a static constraint, or a post-processing adjustment. Our paper instead treats the recommendation list as a control variable. Ranking does not only reveal predicted preferences; it allocates attention and thereby changes the demand process faced by the inventory system.

A closely related stream studies inventory-aware and capacity-aware recommendation. Demirezen and Kumar (2016) show that inventory information can improve recommendation decisions in a subscription-based DVD rental setting. Related work uses inventory, capacity, or context information to adjust recommendation or ranking decisions (Musto et al., 2021; Erginbas et al., 2023). These papers establish that inventory signals matter for recommendation design. We build on this insight but shift the object of analysis from an inventory-adjusted recommendation score to a dynamic ranking policy. In our model, recommendation position is a recurring control that changes demand intensities over time while inventory evolves and replenishment decisions are made endogenously. The relevant object is therefore not simply an inventory-adjusted score at a point in time, but a policy that prices the continuation-value cost of induced demand.

The second related literature studies inventory-constrained demand management in revenue management and operations. A large body of work examines how firms manage scarce inventory or capacity through prices, substitution, assortment choice, stocking, and dynamic allocation. Foundational work in revenue management analyzes dynamic pricing and capacity allocation under stochastic demand (Gallego and Van Ryzin, 1994, 1997; Talluri and Van Ryzin, 2006). Related inventory models study substitution and stockout-based demand reallocation across products (Smith and Agrawal, 2000; Mahajan and Van Ryzin, 2001; Honhon et al., 2010). Other papers jointly optimize pricing, assortment, and learning under demand uncertainty (Bitran and Ferrer, 2007; Dong et al., 2009; Caro and Gallien, 2007; Rusmevichientong and Topaloglu, 2012; Miao and Chao, 2021). This literature provides the economic foundation for our model because it formalizes how scarce inventory creates shadow costs and how demand-shaping decisions should internalize those costs.

A more directly related set of papers combines personalization or assortment decisions with inventory constraints. Ryzin and Mahajan (1999) study the trade-off between variety and inventory risk in retail assortments. Bernstein et al. (2015) develop a dynamic assortment-customization model with limited inventories and state-dependent inventory thresholds. Golrezaei et al. (2014) study real-time personalized assortments under fixed inventories and uncertain choice. Topaloglu (2013) analyzes joint stocking and product-offer decisions under multinomial logit choice. A related algorithmic literature develops approximation and greedy methods for dynamic assortment problems under richer choice models (Aouad et al., 2018, 2019, 2021). More recent work studies inventory-constrained recommendation at checkout and the coordination of stocking with assortment personalization (Chen et al., 2024; Bai et al., 2025). These papers share our interest in the interaction between demand management and inventory constraints. The key distinction is the control variable. They typically choose prices, stocking levels, or the composition of an offered set. We study the ordered sequence of products within a displayed recommendation list. Ordered exposure reallocates attention without changing prices, product availability, or the set of products displayed.

The third related literature studies strategic resource allocation and organizational misalignment in platforms and firms. A common theme in this work is that local decisions or local performance metrics can improve one part of the system while imposing costs elsewhere. In multitask agency models, incentives tied to incomplete performance measures can distort effort away from unmeasured or weakly measured tasks (Holmstrom and Milgrom, 1991; Feltham and Xie, 1994). More recent work studies analogous tensions in platform settings, where demand-generation decisions interact with constrained operating resources. Cachon et al. (2017) study pricing on service platforms with self-scheduling capacity, and Bimpikis et al. (2019) analyze how platform pricing affects spatial supply allocation in ride-sharing networks. Li et al. (2023) study marketplace scalability and show how platform investment interacts with seller capacities and sellers' own investment incentives. Li et al. (2024) study A/B testing with spillovers and show that improvements in a measured primary metric can reduce overall firm performance when they create negative effects on unmeasured dimensions. This organizational logic is close to the misalignment in our setting: a recommendation system can improve relevance, click-through, conversion, or short-run revenue while imposing inventory and fulfillment costs elsewhere in the firm. A similar scarce-resource allocation issue appears in our model through ordered recommendation exposure: a local demand-shaping decision reallocates customer attention toward products whose inventory, backlog, and replenishment states determine the system-wide cost of the induced demand.

Finally, our solution approach belongs to the restless-bandit and index-policy literature. Restless bandits, introduced by Whittle (1988), generalize classical multi-armed bandits by allowing each arm to evolve even when it is not activated. The framework is powerful but computationally difficult: related queueing-control problems are EXP-complete (Papadimitriou and Tsitsiklis, 1999), and general restless-bandit problems are hard to approximate (Guha et al., 2010). Index policies provide tractable priority rules by assigning each arm a scalar value and activating arms with the highest indices (Weber and Weiss, 1990; Niño-Mora, 2001). The Whittle index is asymptotically optimal in several structured environments (Verloop, 2016), but closed-form indices are rare, and many applications rely on numerical marginal-productivity calculations (Niño-Mora, 2007). Multi-mode restless bandits, in which each arm can take more than two activity levels, have been studied by Glazebrook et al. (2006). Our model falls in this class because each product is an arm and recommendation positions are attention modes, with replenishment as an additional action. We derive a closed-form index for the relaxed single-product problem and use it to construct a scalable assignment policy for the coupled ranking problem.

These literatures position our paper along four dimensions. Recommendation models predict what consumers may want, but typically do not price the future inventory cost of the demand they stimulate. Inventory and revenue-management models internalize scarcity, but usually do not treat ordered recommendation exposure as the demand-shaping control. Platform, agency, and organizational resource-allocation models show how local objectives can create system-wide costs, but they do not study ordered recommendation exposure as the mechanism through which such costs arise. Restless-bandit methods provide a tractability framework, but have not, to our knowledge, delivered a closed-form multi-mode index for inventory-sensitive recommendation ranking with recurring replenishment. Our paper combines these elements in a continuous-time model that jointly optimizes recommendation order and product-specific replenishment while accounting for margins, demand potential, holding costs, backlog costs, and the shadow value of inventory.

3. Model and Problem Formulation

We consider a firm that sells N substitutable products through an online channel. Time is continuous, indexed by $t \geq 0$, and future costs are discounted at rate $\beta > 0$. Each product $i \in \{1, \ldots, N\}$ faces stochastic demand that depends on its position in a recommendation list, and the firm dynamically controls both the recommendation order and product-specific replenishment decisions. Our goal is to characterize the joint recommendation and inventory policy that minimizes expected discounted cost (equivalently, maximizes discounted profit).

3.1. **A Microfoundation for Demand Potential and Position-Based Attention.** We use a simple inspection problem to motivate the demand structure used in the model. The construction separates two objects. Product i has an item-level demand-potential parameter $\theta_i \geq 0$, which measures its ability to generate demand conditional on receiving consumer attention. Positions differ in how much attention they receive; these differences are summarized by position weights $\alpha_1, \ldots, \alpha_N$. The inspection problem below is used to derive the monotonicity of these position weights.

The product-specific demand potential θ_i is then introduced as a reduced-form parameter capturing the probability of purchase, which can be thought of as a crude proxy of quality/value/worth depending on the particular context.

Let $\mathcal{N} = \{1, \ldots, N\}$ denote the set of products. Let $\mathcal{P}(N)$ denote the set of rankings, where a ranking is a bijection $R : \mathcal{N} \to \mathcal{N}$. We write $R(i) = k$ when product i is assigned to position k in the recommendation list.

The consumer views the retailer's displayed list as a curated set of relevant alternatives. Inclusion in the line-up is informative: it tells the consumer that the product is available, belongs to the relevant category, is compatible with the query or filter, and is plausibly worth considering. The within-list order, however, is not treated as a credible ranking of the consumer's own idiosyncratic match values. The order may reflect seller-side variables such as margin, inventory, sponsorship, generic popularity, or experimentation. Thus earlier positions are inspected because they are more salient and less costly to compare, not because the consumer believes they necessarily contain the highest personal matches.

If $\ell = (i_1, \ldots, i_N)$ is the ordered displayed list, let $\mathcal{C}(\ell)$ denote the consumer's coarse pre-inspection information. This information records the displayed assortment and its coarse context but does not decode the permutation of positions into residual personal match values. In particular,

$$\mathcal{C}(i_1, \ldots, i_N) = \mathcal{C}(i_{\pi(1)}, \ldots, i_{\pi(N)})$$

for every permutation π. Let

$$\mathcal{H} = \mathcal{C}(\ell)$$

denote the realized coarse information. The retailer may choose the order strategically, but the consumer does not regard that order as a reliable signal of the residual match values defined below.

Let X_j denote the residual match value that the consumer would learn from inspecting the product in position j. This is not objective product quality. It is the part of the consumer's value that remains unresolved after the immediately available information has been processed. Observable product differences that are already salient to the consumer are included in $\mathcal{H}$. The random variables X_j represent the remaining idiosyncratic fit, compatibility, style, use-case fit, or other match components learned only through inspection.

Assumption 1 (Residual exchangeability)**.** *Conditional on $\mathcal{H}$, the residual match values*

$$(X_1, \ldots, X_N)$$

are exchangeable and bounded above.

The assumption is a restriction on the consumer's posterior-predictive beliefs, not a claim that the products are objectively identical. Products may differ in ratings, prices, margins, inventory states, and baseline demand potential. Such differences are either included in the consumer's coarse information $\mathcal{H}$ or summarized in the product-level parameter θ_i. Assumption 1 applies only to the residual match component that is learned through costly inspection.

Consumer heterogeneity gives a natural interpretation. The displayed line-up is a pool of plausible alternatives for consumers with different tastes, shopping missions, and use cases. A consumer may know her broad need, but before inspection, she may not know which listed product will realize the best idiosyncratic fit. If the retailer's within-list order reflects seller-side objectives or population-level considerations rather than the current consumer's residual match, then the consumer treats the location of her best personal match within the curated assortment as symmetric across positions. Selection into the displayed set is informative, while position within that set is not taken to rank residual personal match.

If the consumer inspects the first L positions, she observes $X_1, \ldots, X_L$ and chooses the best inspected alternative. Normalize the outside option to zero and define

$$M_L = \max\{0, X_1, \ldots, X_L\},$$

$$U(L) = \mathbb{E}[M_L \mid \mathcal{H}], \qquad U(0) = 0.$$

Thus $U(L)$ is the gross expected discovery value from inspecting the first L positions. Let

$$\Delta U_k = U(k) - U(k-1) \tag{1}$$

denote the incremental gross value of inspecting position k.

Lemma 1 (Diminishing discovery value)**.** *Under Assumption 1, $U(L)$ is increasing and discretely concave:*

$$\Delta U_k \geq 0, \qquad k = 1, \ldots, N, \tag{2}$$

and

$$\Delta U_{k+1} \leq \Delta U_k, \qquad k = 1, \ldots, N-1.$$

Lemma 1 establishes that returns to discovering more items are "concave". It is natural to think that this discovery process is a costly endeavor for the consumers, varying from individual to individual. We model this by allowing consumers differ in the cost of attention. Let $L \in \{0, 1, \ldots, N\}$ denote the number of positions inspected. Inspecting the first L positions requires attention cost

$$\begin{aligned} \kappa_L &= \sum_{r=1}^{L} c_r, \qquad\qquad \kappa_0 = 0, \\ 0 &< c_1 \leq c_2 \leq \cdots \leq c_N. \end{aligned} \tag{3}$$

The weak increase in c_r captures the fact that lower positions are no easier to inspect than higher positions.

Combining Lemma 1 with (3), the benefit-cost ratios of deeper inspection are weakly decreasing:

$$\frac{\Delta U_1}{c_1} \geq \frac{\Delta U_2}{c_2} \geq \cdots \geq \frac{\Delta U_N}{c_N}. \tag{4}$$

Thus decreasing returns to attention are not imposed directly. They follow from residual exchangeability and weakly increasing marginal inspection costs.

Let $\tau \geq 0$ denote the consumer's attention-cost type, drawn from distribution F_τ. A type-τ consumer pays attention cost $\tau\kappa_L$ when inspecting L positions and chooses

$$L^*(\tau) \in \arg\max_{L\in\{0,1,\dots,N\}} \{U(L) - \tau\kappa_L\}. \tag{5}$$

Ties are broken in favor of the larger inspection depth; this convention is irrelevant when F_τ has no mass at the threshold values below.

Because

$$U(L) - \tau\kappa_L = \sum_{r=1}^{L} (\Delta U_r - \tau c_r),$$

the incremental net benefit from inspecting position k is

$$\Delta U_k - \tau c_k.$$

By (4), the sequence of incremental net benefits crosses zero at most once. Therefore position k is inspected if and only if

$$\tau \leq \frac{\Delta U_k}{c_k}. \tag{6}$$

The probability that position k is inspected is therefore

$$\alpha_k = \Pr(L^*(\tau) \geq k) = F_\tau\left(\frac{\Delta U_k}{c_k}\right). \tag{7}$$

Since $\Delta U_k/c_k$ is weakly decreasing in k, and F_τ is nondecreasing,

$$1 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N \geq 0. \tag{8}$$

In the dynamic model, we treat these α_k's as fixed position weights. This corresponds to conditioning on a maintained recommendation context $\mathcal{H}$, or equivalently to using position weights estimated after averaging over a stable distribution of such contexts.

The inspection problem is used only to derive the probability that each position receives attention. We do not use the residual match values X_j to derive the full product-choice probability within the inspected set. Instead, product-level conversion or request intensity conditional on attention is summarized by the parameter θ_i. This separation permits the demand rate below to be written as a product of a position-attention term and a product-specific demand-potential term.

The position weights therefore arise from consumer optimization under costly attention. Consumers inspect a prefix of the displayed list because earlier positions are more salient and less costly to compare. The residual exchangeability assumption disciplines the value of deeper inspection by making each additional inspected product another opportunity to discover a high residual match from the same posterior-predictive pool. The position- effect and ordered-search literature motivates the prefix-inspection environment (e.g., Varian, 2007; Lahaie, 2008; Athey and Ellison, 2011); the additional behavioral restriction here is that the within-list order is not decoded as a ranking of residual personal match.

We now combine the position-based attention component with product-level demand potential. The inspection problem determines the probability α_k that position k receives attention. Product i's parameter θ_i measures demand potential conditional on receiving attention, with any aggregate arrival-rate scale absorbed into θ_i. Thus, if product i is assigned to position $R(i)$, its induced demand rate is $\alpha_{R(i)}\theta_i$.

Definition 1 (Position-weighted demand)**.** *The position weights satisfy*

$$1 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N \geq 0$$

and $\theta_i \geq 0$ for all i. If ranking $R \in \mathcal{P}(N)$ is in force, the demand rate for product i is

$$\lambda_i(R) = \alpha_{R(i)}\theta_i. \tag{9}$$

Equation (9) is the demand structure used in the main model. It decomposes demand into a product-specific component, θ_i, and a position-specific attention component, $\alpha_{R(i)}$. The first component does not depend on the ranking. The second component does not depend on the identity of the product assigned to that position.

This decomposition makes recommendation ranking an allocation of scarce attention across heterogeneous products. In a static environment, high-attention positions would naturally be assigned to products with large θ_i. With dynamic inventory, however, demand potential alone is not the relevant ranking criterion. The firm must also account for the inventory cost of generating additional demand for each product.

3.2. **Inventory States and Actions.** Let

$$S_i = \{m \in \mathbb{Z} : \underline{K}_i \le m \le K_i\}, \qquad \underline{K}_i < 0 < K_i,$$

denote the feasible inventory levels for product i. Positive values represent on-hand stock, while negative values represent backlog. Backlog is capped at $\underline{K}_i$ to keep the state space finite. The joint inventory state is

$$\boldsymbol{s} = (s_1, \ldots, s_N) \in S := \prod_{i=1}^{N} S_i.$$

The primitive control problem has two decision components. The first is a gradual ranking control. Between replenishment interventions, the firm maintains a ranking

$$R \in \mathcal{P}(N),$$

where $\mathcal{P}(N)$ is the set of permutations of $\{1, \ldots, N\}$. We write $R(i) = k$ when product i is assigned to position k in the recommendation list.

The second component is replenishment. For each product i, define the reset map

$$\Gamma_i(\boldsymbol{s}) = (s_1, \ldots, s_{i-1}, K_i, s_{i+1}, \ldots, s_N).$$

A replenishment intervention for product i returns the inventory state from $\boldsymbol{s}$ to $\Gamma_i(\boldsymbol{s})$.

For dynamic programming, we work with the standard CTMDP representation in which the primitive gradual-impulse problem is translated into a CTMDP with gradual controls only; see Piunovskiy and Zhang (2020, Chapter 7). Thus both rankings and reset interventions are represented as controls in a common finite action set,

$$A = \mathcal{P}(N) \cup \{\mathcal{R}_1, \ldots, \mathcal{R}_N\}.$$

A ranking action $R \in \mathcal{P}(N)$ is held over calendar time and allocates recommendation attention across products. A reset action $\mathcal{R}_i$ is the gradual-control translation of the product-i replenishment intervention. The notation collects both types of decisions in the same CTMDP action set, but it does not identify a reset with a recommendation ranking.

It is without loss to use product-specific reset controls. If the primitive policy calls for several products to be reset at the same intervention date, the same post-intervention state is obtained by applying the corresponding single-product reset maps in any order. The the costs are additive, and the ordered representation differs from the simultaneous intervention only at the intervention instant which has measure zero. Thus the one-reset-at-a-time convention is a convenient representation of the impulse component.

3.3. **Inventory Dynamics and CTMC Generator.** Under a ranking action R, sales of product i arrive according to a Poisson process with rate

$$\lambda_i(\boldsymbol{s}, R) = \alpha_{R(i)}\theta_i \mathbf{1}\{s_i > \underline{K}_i\}.$$

Each sale reduces product i's inventory by one unit. If $s_i = \underline{K}_i$, further demand for product i is lost or redirected and does not change the state.

Thus, under ranking R, the nonzero off-diagonal transition rates are

$$q(\boldsymbol{s} - \boldsymbol{e}_i \mid \boldsymbol{s}; R) = \lambda_i(\boldsymbol{s}, R), \qquad i = 1, \dots, N,$$

where $\boldsymbol{e}_i$ is the ith unit vector in $\mathbb{Z}^N$. The diagonal entry is

$$q(\boldsymbol{s} \mid \boldsymbol{s}; R) = -\sum_{i=1}^{N} \lambda_i(\boldsymbol{s}, R).$$

Under the translated reset control $\mathcal{R}_i$, the state moves deterministically to $\Gamma_i(\boldsymbol{s})$. For states with $s_i < K_i$, the normalized transition rates are

$$q(\Gamma_i(\boldsymbol{s}) \mid \boldsymbol{s}; \mathcal{R}_i) = 1, \qquad q(\boldsymbol{s} \mid \boldsymbol{s}; \mathcal{R}_i) = -1,$$

with all other off-diagonal rates equal to zero. The unit rate is part of the gradual CTMDP translation of the replenishment impulse. It is not a physical replenishment lead time. At states with $s_i = K_i$, the reset control has no state effect and is omitted from the feasible action set.

3.4. **Costs and Objective.** The firm incurs holding and backlog costs and earns a margin from each sale. Let $h_i > 0$ denote the per-unit holding cost for product i, and let $p_i > h_i$ denote the per-unit backlog penalty. The instantaneous inventory cost at state $\boldsymbol{s}$ is

$$c(\boldsymbol{s}) = \sum_{i=1}^{N} \left[h_i (s_i)_+ + p_i(-s_i)_+ \right],$$

where $(x)_+ = \max\{x, 0\}$.

Each fulfilled sale of product i generates margin $m_i \geq 0$, which we treat as a negative cost. Under ranking R, the flow cost rate is

$$C(\boldsymbol{s}, R) = \sum_{i=1}^{N} \ell_i(s_i, R), \tag{10}$$

where

$$\ell_i(s_i, R) = h_i(s_i)_+ + p_i(-s_i)_+ - m_i \lambda_i(\boldsymbol{s}, R).$$

For reset controls, $C(\boldsymbol{s}, \mathcal{R}_i)$ is the gradual CTMDP translation of the primitive replenishment impulse cost. With the normalized reset rate introduced above, we write

$$C(\boldsymbol{s}, \mathcal{R}_i) = T_i + b_i(K_i - s_i).$$

This term is not a sales, holding, or backlog flow. It is the cost rate attached to the translated reset control.

Let Π denote the set of admissible nonanticipative policies for the translated CTMDP. Given an initial distribution γ over S, the discounted cost of policy π is

$$J^\pi(\gamma) = \mathbb{E}_\gamma^\pi \left[\int_0^\infty e^{-\beta t} C(\boldsymbol{S}_t, A_t)\, dt \right].$$

The value function is

$$v(\boldsymbol{s}) = \inf_{\pi \in \Pi} J^\pi(\delta_{\boldsymbol{s}}).$$

3.5. **CTMDP Formulation and Complexity.** The value function v satisfies the continuous-time Bellman equation

$$\beta v(\boldsymbol{s}) = \min_{a \in A} \left\{ C(\boldsymbol{s}, a) + \sum_{\boldsymbol{s}' \in S} q(\boldsymbol{s}' \mid \boldsymbol{s}; a)\, v(\boldsymbol{s}') \right\}, \tag{11}$$
$$\boldsymbol{s} \in S.$$

While theoretically well-posed, solving (11) is computationally intractable for realistic catalogs. The state space size $|S| \approx K^N$ grows exponentially in N. Furthermore, discrete-time analogues of this problem are known to be PSPACE-hard (Papadimitriou and Tsitsiklis, 1999). This motivates the development of a tractable index policy.

4. Index Policy via Restless-Bandit Relaxation

In this section, we develop a computationally efficient policy by relaxing the global constraint that couples products through the ranking decision. We formulate the model as a multi-mode restless-bandit problem, in which products are arms and recommendation positions are attention modes. We then apply a Whittle-style Lagrangian relaxation and derive a closed-form index representation of the marginal value of assigning recommendation attention to a product in a given inventory state.

4.1. **Multi-Mode Restless-Bandit Representation and Relaxation.**

We now rewrite the translated CTMDP in product-level terms. Each product is a restless arm. The ranking positions are its operating modes: assigning product i to position $a \in \{1, \ldots, N\}$ gives it demand rate $\alpha_a \theta_i$. The original system-level constraint is that each position is assigned to exactly one product:

$$\sum_{i=1}^{N} \mathbf{1}\{a_i(t) = k\} = 1, \qquad k = 1, \ldots, N.$$

The reset control remains product-specific. For a single product, the local control set is

$$\mathcal{A}_i = \{1, \ldots, N, \mathcal{R}\}.$$

Actions $1, \ldots, N$ are ranking modes, while $\mathcal{R}$ is the translated reset control.

The ranking modes consume recommendation attention. Since α_a is the position-specific attention factor in the demand rate, we use it as the resource consumption of position a:

$$r(a) = \alpha_a, \qquad a = 1, \ldots, N.$$

The translated reset control consumes no recommendation attention, so

$$r(\mathcal{R}) = 0.$$

Thus the Lagrange multiplier introduced below is a shadow price of discounted attention. The integer labels $a = 1, \ldots, N$ continue to encode the ranking assignment constraint, while $r(a)$ measures the attention resource priced in the relaxed single-product problem.

Following Whittle (1988), we relax the one-product-per-position constraint by charging a Lagrange multiplier η for discounted attention. Suppressing the product index, the resulting single-product problem is

$$\min_{\pi} \ \mathbb{E}^{\pi} \int_0^{\infty} e^{-\beta t} \{C(s_t, a_t) + \eta r(a_t)\} \, dt.$$

Here $a_t \in \{1, \ldots, N, \mathcal{R}\}$.

The scalar η is the shadow price of recommendation attention. The relaxation separates the global problem into independent single-product problems. After the product-level attention indices are computed, the original ranking constraint is restored by solving a static assignment problem over products and positions.

4.2. **Closed-form Index Policy.** We now state the single-product attention index used to construct the ranking policy. Throughout this subsection, product subscripts are suppressed. Thus s denotes the inventory state of one product, K its upper inventory level, θ its demand-potential parameter, m its margin, h and p its holding and backlog costs, and T and b its fixed and unit replenishment costs. A ranking action $a \in \{1, \ldots, N\}$ assigns the product to position a. Smaller values of a correspond to higher attention. The demand rate in position a is $\alpha_a \theta$, and the attention consumed by that position is$r(a) = \alpha_a$.

The relaxation introduced above prices attention by a scalar charge $\eta \geq 0$. For a fixed single-product it is well known that we can limit attention to Markov policies, denoted generically as S. Thus η is the shadow price of one unit of discounted attention. The index is the critical value of this charge at which the product is indifferent between local attention levels.

The calculations below are written for a fixed continuation policy S. It is easy to see that in the single arm problem if it is optimal to reset/replenish at some level s it is also optimal to do so at any lower level. Let $\bar{s}$ denote the largest state at which S uses reset. Then states $s \leq \bar{s}$ are reset states, while states $s > \bar{s}$ are ranking states. On the ranking region, S assigns a position $a_S(x) \in \{1, \dots, N\}$ to each active inventory state x. The monotonicity result in Appendix B.1 directly implies that in the optimal policy lower inventory states receives weakly lower attention: if $s_1 < s_2$ and both states are in the ranking region, then

$$a_S(s_1) \geq a_S(s_2).$$

That is, in the optimal policy we can identify a *partition* of the state space where at the bottom is a replenishment region and each higher partition has "higher" actions.[2] For a fixed partition continuation policy S, define the demand rate induced at an active state $x > \bar{s}$ by

$$\lambda_x = \alpha_{a_S(x)}\theta.$$

For a ranking action a, the single-product flow cost and attention are

$$\begin{aligned} C(x,a) &= h(x)_+ + p(-x)_+ - \alpha_a \theta m, \\ r(a) &= \alpha_a. \end{aligned} \tag{12}$$

Therefore, under S,

$$\begin{aligned} c_S(x) &= C\big(x, a_S(x)\big), \\ r_S(x) &= r\big(a_S(x)\big), \qquad x > \bar{s}. \end{aligned} \tag{13}$$

[2] With the convention that ranking 1, is treated as the highest action whereas ranking N is the lowest.

At the reset boundary, the product is replenished to K through the translated unit-rate reset control, so

$$c_S(\bar{s}) = T + b(K - \bar{s}), \qquad r_S(\bar{s}) = 0. \tag{14}$$

Let X_t be the single-product inventory process under S. Define

$$\begin{aligned} F_S(x) &= \mathbb{E}_x^S\left[\int_0^\infty e^{-\beta t} c_S(X_t)\,dt\right], \\ G_S(x) &= \mathbb{E}_x^S\left[\int_0^\infty e^{-\beta t} r_S(X_t)\,dt\right]. \end{aligned} \tag{15}$$

The charged value of a policy S is

$$F_S(x) + \eta G_S(x).$$

Because sales are happening one at a time for every active state $x > \bar{s}$, F_S and G_S satisfy

$$\begin{aligned} F_S(x) &= \frac{c_S(x) + \lambda_x F_S(x-1)}{\beta + \lambda_x}, \\ G_S(x) &= \frac{r_S(x) + \lambda_x G_S(x-1)}{\beta + \lambda_x}. \end{aligned} \tag{16}$$

The reset boundary conditions are

$$\begin{aligned} F_S(\bar{s}) &= \frac{T + b(K - \bar{s}) + F_S(K)}{\beta + 1}, \\ G_S(\bar{s}) &= \frac{G_S(K)}{\beta + 1}. \end{aligned} \tag{17}$$

Equations (16)–(17) show how the primitive economic objects enter the index. Holding and backlog costs enter through $h(x)_+ + p(-x)_+$, margins enter through $-\alpha_a \theta m$, replenishment costs enter through $T + b(K - \bar{s})$, and position-dependent demand enters through $\alpha_a \theta$.

The index is obtained from a one-shot deviation from the partition policy S. Fix an active state $s > \bar{s}$, and suppose that the background policy assigns position

$$a = a_S(s).$$

For an alternative position $k \in \{1, \ldots, N\}$, let $F_{(k,S)}(s)$ and $G_{(k,S)}(s)$ denote the discounted operating cost and discounted attention obtained by assigning position k only until the first transition out of state s, and then following S thereafter. The first-exit values are

$$F_{(k,S)}(s) = \frac{C(s,k) + \alpha_k \theta F_S(s-1)}{\beta + \alpha_k \theta},$$
$$G_{(k,S)}(s) = \frac{\alpha_k + \alpha_k \theta G_S(s-1)}{\beta + \alpha_k \theta}. \tag{18}$$

At charge η, the critical comparison is

$$F_S(s) + \eta G_S(s) = F_{(k,S)}(s) + \eta G_{(k,S)}(s).$$

Solving gives

$$\nu(k, s; S) = \frac{F_{(k,S)}(s) - F_S(s)}{G_S(s) - G_{(k,S)}(s)} \tag{19}$$

whenever the denominator is nonzero. This is the usual marginal-productivity ratio: the numerator is the marginal operating-cost change and the denominator is the marginal discounted attention change (Nino-Mora, 2006; Niño-Mora, 2007).

The attention normalization makes the adjacent denominator positive. If $k = a+1$ and $\alpha_a > \alpha_{a+1}$, then $G_S(s) - G_{(a+1,S)}(s) > 0$. If $\alpha_a = \alpha_{a+1}$, the two adjacent positions are attention-equivalent and can be tied.

Define

$$\Delta F_S(s) := F_S(s) - F_S(s-1),$$

and

$$\Delta G_S(s) := G_S(s) - G_S(s-1).$$

Proposition 1 (Closed-form attention index)**.** *Fix a single-product partition policy S, and consider an active state $s > \bar{s}$. If the background position $a = a_S(s)$ satisfies $\alpha_a \neq \alpha_k$, then the one-shot critical charge for switching from a to any position $k \neq a$ is*

$$\psi(s; S) = \frac{\theta\big[m + \Delta F_S(s)\big]}{1 - \theta\, \Delta G_S(s)}. \tag{20}$$

In particular, the critical charge is independent of the deviation.

The denominator in (20) is positive as we show in Appendix B.3. Alternatively, write

$$c(s) = h(s)_+ + p(-s)_+.$$

Then

$$\psi(s; S) = \frac{\theta\{c(s) + \beta m - \beta F_S(s-1)\}}{\beta\{1 + \theta G_S(s-1)\}}. \tag{21}$$

A compact occupation-measure equivalent follows from the same recursions which is convenient for numerical calculations is derived in Appendix B.3.

The numerator of (20) has a direct inventory-value interpretation. One unit of attention generates demand at rate θ. A demand arrival in state s moves inventory from s to $s-1$. The term m is the immediate margin from the sale, while $\Delta F_S(s)$ is the continuation-cost effect of losing one unit of inventory under the background policy. Thus

$$m + \Delta F_S(s)$$

is the inventory-adjusted marginal value of demand, and $\theta[m + \Delta F_S(s)]$ is the value of an additional unit of attention.

The denominator adjusts for the induced change in future attention. Hence $\psi(s; S)$ is the shadow value of discounted recommendation attention at inventory state s. A larger value means that the product can bear a higher attention charge before it is optimal to reduce its assigned attention.

4.3. **Optimal Partition and Ranking Assignments.** Fix an attention charge η. A charge-optimal single-product policy is identified by charged indifference conditions among ranking actions and by the analogous indifference condition between ranking and reset. The closed-form index derived above does not replace these indifference conditions. It shows that, under the attention normalization $r(a) = \alpha_a$, all ranking-to-ranking critical charges at a given state collapse to the same state-level attention value.

Consider an active state $s > \bar{s}$ at which the continuation policy assigns position

$$a = a_S(s).$$

For any alternative ranking position $k \in \{1, \ldots, N\}$, the switching condition between the background action a and the one-shot action k is

$$F_S(s) + \eta G_S(s) = F_{(k,S)}(s) + \eta G_{(k,S)}(s). \tag{22}$$

When $\alpha_k \neq \alpha_a$, Proposition 1 gives

$$\nu(k, s; S) = \psi(s; S).$$

Thus (22) is equivalently

$$\eta = \psi(s; S). \tag{23}$$

If $\alpha_k = \alpha_a$, the two positions are attention-equivalent and can be tied.

Therefore the active-state partition is still identified by charged indifference, but the deviation (k) independence of the index makes it remarkably simple: the relevant object is the product-state shadow value $\psi(s; S)$. Appendix B.4 shows that, on each sign-homogeneous block, the switching equation

$$\eta = \psi(s; S)$$

reduces to an affine equation in a geometric term and therefore has at most one solution, except in the degenerate case in which the expression is constant on the whole block. Collecting these switching states yields the monotone active-state partition used in the single-product calculation.

The replenishment boundary is determined by the same charged-indifference logic. If the product is reset at state s, the first-exit values are

$$\begin{aligned} F_{(\mathcal{R},S)}(s) &= \frac{T + b(K - s) + F_S(K)}{\beta + 1}, \\ G_{(\mathcal{R},S)}(s) &= \frac{G_S(K)}{\beta + 1}. \end{aligned} \tag{24}$$

The ranking-versus-reset boundary is characterized by

$$F_S(s) + \eta G_S(s) = F_{(\mathcal{R},S)}(s) + \eta G_{(\mathcal{R},S)}(s). \tag{25}$$

Equivalently, whenever the denominator is nonzero, the reset critical charge is

$$\nu(\mathcal{R}, s; S) = \frac{F_{(\mathcal{R},S)}(s) - F_S(s)}{G_S(s) - G_{(\mathcal{R},S)}(s)}. \tag{26}$$

The reset boundary solves the corresponding indifference equation

$$\eta = \nu(\mathcal{R}, s; S).$$

Since the reset region is downward closed, the replenishment rule is summarized by the largest reset state $\bar{s}$.

The scalar η is not an additional inventory state. It is the dual price of discounted attention from the Lagrangian relaxation. The single-product calculation uses η to identify the active-state partition through $\eta = \psi(s; S)$ and the replenishment boundary through (25). The hard-constrained recommendation problem then assigns physical ranking positions across products.

For each product i, let S_i^* denote the single-product continuation policy obtained from these charged switching equations, and let $\bar{s}_i$ be its reset boundary. At a global state $\boldsymbol{s} = (s_1, \ldots, s_N)$, define

$$\mathcal{J}(\boldsymbol{s}) := \{i : \; s_i \le \bar{s}_i\}, \qquad \mathcal{I}(\boldsymbol{s}) := \{i : \; s_i > \bar{s}_i\}.$$

Products in $\mathcal{J}(\boldsymbol{s})$ are reset. Products in $\mathcal{I}(\boldsymbol{s})$ participate in the ranking assignment.

For an active product $i \in \mathcal{I}(\boldsymbol{s})$, define its attention index by

$$\psi_i(s_i) = \psi_i(s_i; S_i^*),$$

computed from (20) using product i's primitives and continuation policy S_i^*. The value $\psi_i(s_i)$ is the product-state shadow value of one unit of discounted attention. Assigning product i to position k supplies attention α_k, so the product–position score is

$$\alpha_k \psi_i(s_i).$$

The ranking used by the index policy is obtained by solving

$$\max_{x_{ik} \in \{0,1\}} \sum_{i \in \mathcal{I}(\boldsymbol{s})} \psi_i(s_i) \sum_{k=1}^{N} \alpha_k x_{ik}$$

subject to

$$\sum_{k=1}^{N} x_{ik} = 1 \quad \forall i \in \mathcal{I}(\boldsymbol{s}),$$

and

$$\sum_{i \in \mathcal{I}(\boldsymbol{s})} x_{ik} \leq 1 \quad \forall k.$$

Because $\alpha_1 \geq \cdots \geq \alpha_N$, this assignment ranks active products by decreasing $\psi_i(s_i)$ and assigns higher-attention positions to higher-index products. The assignment formulation is retained because it also handles ties, reset states, inactive products, and variants with fewer visible positions.

This is a maximum-weight assignment problem and is solvable in $O(N^3)$. When the number of visible positions is much smaller than N, the same logic can be implemented by evaluating the product-level attention indices and selecting the highest-index products by partial sort in $O(N \log N)$ time (see Section 5.3).

5. Computational Experiments

This section evaluates the proposed attention-index policy in two environments. We first study a four-product system for which the exact discounted CTMDP can be solved by policy iteration. This benchmark allows us to compare the index policy with the exact optimum, decompose the index–DP gap, and identify where the product-level approximation performs well and where full joint-state coordination remains valuable. We then turn to large systems, where exact dynamic programming is infeasible and the relevant comparison is against scalable heuristics. Throughout, the objective is discounted cost; lower values are better.

5.1. **Experimental Setup.** Unless stated otherwise, we consider four products indexed by $i \in \{1, \ldots, 4\}$. Product 1 is a high-margin, low-demand item; product 2 is a low-margin, higher-demand item; and products 3 and 4 create intermediate and high-penalty cases. The base demand-potential parameter for product i is θ_i, each sale earns margin m_i, and the instantaneous inventory cost is

$$c(\boldsymbol{s}) = \sum_{i=1}^{4} \left[h_i(s_i)_+ + p_i(-s_i)_+ \right],$$

where $h_i > 0$ is the holding cost and $p_i > h_i$ is the backorder penalty. Resetting product i from s_i to its upper bound K incurs fixed cost T_i plus per-unit replenishment cost $b_i(K - s_i)$. Position-attention weights decline in position number; we use

$$\alpha = (1, 0.7, 0.5, 0.3).$$

All products share inventory bounds $s_i \in [K_{\text{lower}}, K]$, with $K = 7$ and $K_{\text{lower}} = -4$.

We compare five policies.

- **Exact DP:** stationary optimal policy obtained by solving the discounted CTMDP[3] using policy iteration.
- **Index:** for each product, we compute its attention index as a function of inventory and use the assignment rule from Section 4. The index is evaluated in closed form under a product-specific continuation policy S_i^*, taken to be the charge-optimal single-product policy at $\eta = 0$; because each relaxed single-product problem has only $K - K_{\text{lower}} + 1$ inventory states, this policy is computed exactly, and it supplies the replenishment boundary $\bar{s}_i$ and the monotone ranking blocks that determine F_S and G_S in the index. Products inside the replenishment region ($s_i \le \bar{s}_i$) receive the lowest ranking priority; in the showcase instance, $\bar{s}_i = -1$ for all four products. The score from assigning product i to position k is $\alpha_k \psi_i(s_i)$. Ranking is obtained by solving the induced assignment problem. We compare ranking actions with product-specific replenishments using a one-step lookahead that evaluates each replenishment against a proxy value function initialized from the demand-only policy.
- **Bootstrapped index:** we refine the index policy's replenishment decisions by iteratively improving the proxy value function used in the rank-versus-replenishment comparison. Starting from the demand-only proxy, we evaluate

[3]We discretize time via uniformization; see Piunovskiy and Zhang (2020) for a textbook treatment.

the resulting index policy under the CTMDP transition structure, use its value function as the new proxy, and repeat. Ranking decisions are unchanged and are determined by the attention index ψ; only the replenishment comparison is refined. In all experiments, the procedure converges in three to five iterations.

- **Demand-only:** a static ranking that ignores inventory, sorting products by $\theta_i m_i$ in descending order. Inventory control uses a simple reorder rule: whenever any product reaches nonpositive inventory, the most backordered product is replenished first.
- **Myopic one-step:** at each state $\boldsymbol{s}$, the policy chooses the action, either ranking or replenishment, that minimizes the current flow cost $C(\boldsymbol{s}, a)$, ignoring future consequences.

We report three performance metrics. The first is discounted total cost $J^{\pi}(\boldsymbol{s}_0)$ from an initial state $\boldsymbol{s}_0$, computed by exact policy evaluation using the linear system induced by the policy's transition matrix and reward structure. The second is discounted backlog share, defined as the fraction of discounted time for which $\min_i s_i < 0$. The third is discounted scarcity share, defined as the fraction of discounted time for which $\min_i s_i \leq 1$. The two service metrics are estimated by Monte Carlo simulation with 1,000 sample paths and horizon $t_{\max} = 8/\beta = 40$, so that the remaining discount mass beyond $t_{\max}$ is negligible.

5.2. **Small System: Exact DP Benchmark.** We begin with a four-product instance that is small enough to solve by exact dynamic programming but rich enough to make the ranking–inventory trade-off nontrivial. The products differ in demand intensity, margin, holding cost, backlog penalty, and replenishment cost, as shown in Table 1. This heterogeneity is deliberate: a demand-only ranking, a margin-based ranking, and an inventory-aware ranking make different choices only when products differ along multiple economic dimensions.

We set the discount rate to $\beta = 0.2$ and choose backorder and replenishment costs large enough to make ranking and replenishment decisions economically consequential. With $K = 7$ and $K_{\text{lower}} = -4$, each product has twelve inventory states, including

TABLE 1. Showcase four-product instance (all items share $K = 7$ and $K_{\text{lower}} = -4$).

Item	θ_i	m_i	h_i	p_i	T_i	b_i
1 (luxury)	1.5	8.0	0.20	32	32	4
2 (commodity)	2.5	1.5	0.10	8	12	2
3 (intermediate)	2.5	3.0	0.15	24	20	3
4 (high-penalty)	1.8	4.0	0.10	64	24	4

positive inventory states, low-stock states, and backlog states. The resulting CTMDP has

$$|\mathcal{S}| = (K - K_{\text{lower}} + 1)^4 = 12^4 = 20{,}736$$

states and

$$|\mathcal{A}| = 4! + 4 = 28$$

actions per state: all potential product rankings and replenishment for each product. Policy iteration converges in eight iterations.

The index policy has two decision margins. The first is the product–position assignment generated by the attention indices. The second is the rank-versus-replenishment comparison that determines when a product should be replenished rather than displayed. This structure motivates a gap decomposition. Let V^{DP}, V^{idx}, and V^{dpR} denote the value functions of the exact DP, the index policy, and a hybrid policy that uses the attention index for ranking but optimizes replenishment timing by solving a reduced dynamic program over replenishment actions. Holding the ranking rule fixed and optimizing only the replenishment margin gives

$$\begin{aligned}\underbrace{V^{\text{idx}}(\boldsymbol{s}) - V^{\text{DP}}(\boldsymbol{s})}_{\substack{\text{total}\\ \text{gap}}} &= \underbrace{V^{\text{idx}}(\boldsymbol{s}) - V^{\text{dpR}}(\boldsymbol{s})}_{\substack{\text{replenishment timing}\\ \text{gap}}} \\ &\quad + \underbrace{V^{\text{dpR}}(\boldsymbol{s}) - V^{\text{DP}}(\boldsymbol{s})}_{\substack{\text{ranking}\\ \text{gap}}}.\end{aligned}$$

The first term is the cost of the index policy's replenishment approximation, conditional on the index ranking rule. The second term is the remaining cost of using product-level ranking indices rather than the state-dependent DP-optimal ranking.

TABLE 2. Gap decomposition at representative inventory states (percentage gap relative to $|V^{\mathrm{DP}}(\boldsymbol{s})|$). The total gap decomposes into a replenishment timing component and a ranking component.

State	$\boldsymbol{s}$	V^{DP}	Total gap	replenishment gap	Ranking gap
Full stock	$(7,7,7,7)$	−36.55	6.3%	4.4%	1.9%
Mid stock	$(3,3,3,3)$	8.19	48.4%	32.8%	15.5%
Low stock	$(1,1,1,1)$	43.46	10.1%	6.3%	3.7%
Mixed	$(7,0,3,-3)$	47.61	3.8%	2.9%	0.9%
Backlog	$(-2,-2,-2,-2)$	75.00	0.0%	0.0%	0.0%
Global average (states with $\lvert V^{\mathrm{DP}}(\boldsymbol{s})\rvert > 1$)			16.7%	13.7%	3.0%

Table 2 reports the decomposition at representative inventory states. Two patterns stand out. First, the index is close to the exact optimum across the board: the total gap is single-digit at full stock (6.3%) and in the mixed state (3.8%), modest at low stock (10.1%), and exactly zero in deep backlog. Second, the ranking component—the irreducible cost of selecting rankings through independent product-level indices—is small everywhere, never exceeding 3.7% except at mid stock. The apparent spike at mid stock (total 48.4%) is a small-denominator artifact: $|V^{\mathrm{DP}}| = 8.19$ is by far the smallest value among the diagnostic states, so a modest absolute difference (here $V^{\mathrm{idx}} - V^{\mathrm{DP}} \approx 4.0$, comparable to the 2.3 absolute difference at full stock) translates into a large percentage. In absolute terms the index tracks the exact optimum comparably well at full and mid stock. The economics are intuitive: when inventories are scarce or backlogged, the dominant trade-off is local—whether additional exposure to a product is worth the inventory cost it creates—and this is precisely the marginal gain captured by the attention index. The residual gap reflects the cross-product coordination available only to full joint-state DP, which matters most at intermediate inventory.

The bootstrapped index eliminates the replenishment timing component by iteratively improving the proxy value function used in the rank-versus-replenishment comparison. Across all 20,736 states, the bootstrapped index matches the DP-optimal-replenishment hybrid and reduces the global average gap (computed over states with $|V^{\mathrm{DP}}| > 1$) from 16.7% to 3.0%. The remaining gap is the ranking component:

Table 3. Performance in the showcase instance. Lower is better for all metrics. J is computed via exact policy evaluation; service metrics via 1,000-run Monte Carlo simulation.

Initial state	Metric	DP	Index	Boot. Idx	Myopic	Demand
`full`	J	−36.55	−34.24	−35.85	−21.33	−17.50
	Backlog share	0.280	0.329	0.264	0.471	0.000
	Scarcity share	0.507	0.509	0.504	0.585	0.414
`scarce1`	J	−5.79	−2.84	−4.90	26.47	23.96
	Backlog share	0.367	0.436	0.355	0.778	0.000
	Scarcity share	0.930	0.930	0.925	1.000	0.596
`mixed`	J	1.41	5.79	1.86	31.69	29.02
	Backlog share	0.393	0.519	0.386	0.761	0.000
	Scarcity share	0.932	0.925	0.931	1.000	0.760
`backlog`	J	22.45	29.36	22.88	74.19	47.29
	Backlog share	0.468	0.635	0.462	1.000	0.171
	Scarcity share	0.824	0.776	0.824	1.000	0.592
Weighted avg.	J	1.77	6.27	2.37	37.57	28.33
	Backlog share	0.396	0.510	0.387	0.809	0.051
	Scarcity share	0.857	0.840	0.854	0.959	0.626

the cost of selecting rankings through independent product-level indices rather than through full joint-state coordination. This residual is small in absolute terms and, in the scarcity states that drive operating performance, close to zero.

We next compare the exact DP, the base index, the bootstrapped index, and the two implementable baselines at four diagnostic initial states: full inventory $(7, 7, 7, 7)$, one scarce product $(1, 7, 7, 7)$, mixed inventory $(7, 1, 7, 1)$, and a backlog state $(7, -1, 7, 0)$. We also report a weighted average that places 10% weight on the full-inventory state and 30% on each of the remaining states. Table 3 reports discounted total cost from exact policy evaluation and the two service metrics from Monte Carlo simulation.

Table 3 gives three main findings. First, the bootstrapped index is the best-performing implementable policy on weighted-average cost. It reduces cost by 62% relative to the base index, by 92% relative to demand-only ranking, and by 94% relative to myopic control. It also comes close to the exact DP optimum (weighted-average J of 2.37 versus 1.77); the residual gap reflects the cross-product coordination available only to full joint-state DP, whereas the index decomposes the problem into product-level priority calculations. The decomposition in Table 2 confirms that this residual loss is modest at every diagnostic state and vanishes in backlog.

Second, myopic control performs relatively well at full stock but deteriorates sharply when inventory becomes scarce. This is the expected failure mode of a one-step rule: it prices current flow cost but not the future inventory state created by current exposure and replenishment decisions. Demand-only ranking fails for the opposite reason. It avoids backlog aggressively, but its conservative replenishment rule is costly in the discounted objective. Its low backlog share is therefore not evidence of superior performance; it reflects overprotection against backlog rather than an efficient ranking–replenishment trade-off.

Third, the bootstrapped index makes ranking–replenishment trade-offs that are closer to exact DP than either baseline. Its weighted-average backlog and scarcity shares (0.387 and 0.854) are close to the DP values (0.396 and 0.857), while its discounted cost is far below the implementable heuristics. The policy therefore does not simply minimize backlog or maximize short-run revenue. It balances margin, scarcity, and replenishment in a way that is consistent with the inventory-adjusted exposure logic of the model.

The exact DP remains the benchmark and achieves the lowest cost at every diagnostic state. The contribution of the index is different: it delivers a transparent and scalable policy that comes close to the exact optimum and substantially improves on implementable static and myopic rules, while preserving the economic structure of the ranking–replenishment trade-off. We now turn to large systems, where exact DP is infeasible and only scalable policies can be compared.

5.3. **Large Systems: Ranking-Aligned Simulations.** The four-product benchmark highlights the core computational tension. Exact dynamic programming is useful for validation, but it cannot scale to realistic catalogs. We therefore evaluate the index in large systems where firms must rely on scalable policies.

A design issue arises when simulating replenishment at large scale. In the CT-MDP of Section 3, the replenishment action $\mathcal{R}_i$ is modeled as a foreground action with a unit-rate jump. This abstraction keeps the dynamic program tractable while preserving the key economic role of replenishment: resetting inventory changes the future value of exposure. Models with explicit lead times could be incorporated by augmenting the state space with on-order inventory or outstanding replenishment states, but doing so would increase computational complexity. We therefore evaluate two complementary regimes that isolate the ranking–replenishment trade-off without adding lead-time state variables.

- **Regime A (finite-horizon ranking only).** Replenishment actions are switched off. Items deplete toward K_{lower}, and the recommendation list is the only control. This is the ranking sub-chain on which the index is derived, with no auxiliary replenishment mechanism.
- **Regime B (instantaneous replenishment).** replenishment actions are retained but executed immediately. When a product's inventory first reaches $s_j \leq 0$, the firm pays the replenishment cost $T_j + b_j(K - s_j)$ and returns s_j to K in the same decision epoch. There is no lead time and no on-order state, so ranking and replenishment do not overlap.

Across both regimes, we use the same four product types from Table 1, with the same cost primitives but wider inventory bounds, $K = 15$ and $K_{\text{lower}} = -7$, suited to the longer depletion horizons at scale (the tight-inventory stress scenario below uses $K = 10$ and $K_{\text{lower}} = -5$). We construct large catalogs by replicating these types in equal proportions. These simulations are therefore mean-field environments with many copies of a few heterogeneous product classes rather than catalogs with fully heterogeneous items. The index policy is computed once offline as a lookup table that maps each type–inventory pair to a priority score. Online, each decision epoch selects the visible set by choosing products with the highest priority and ordering them accordingly. We compare against two standard implementable heuristics: demand-weighted ranking, which orders products by $\theta_j m_j$, and margin-only ranking, which orders products by m_j. Under the replicated-type construction, these two baselines select the same displayed set for any inventory vector, so we report a single baseline.

To avoid artificial rotation among identical products, we add a small, fixed item-specific tie-breaker to each score, interpretable as persistent idiosyncratic appeal. In both regimes, we simulate up to $t_{\max} = 8/\beta$, use 100 Monte Carlo replications per configuration, and report discounted cost per item J/N, normalized by the discounted time mass $\int_0^{t_{\max}} e^{-\beta t}\,dt$ so that values are expressed in equivalent cost-rate units.

Regime A: Finite-Horizon Ranking Without Replenishment. Table 4 and Figure 1 report discounted cost per item when replenishment actions are turned off. This regime isolates the ranking sub-chain on which the index is derived. The lookup policy dominates the demand-weighted baseline in every scenario. The gains are largest when demand pressure is high and poor ranking quickly pushes high-demand products into backlog. Cost reductions range from 1.9% at $N = 1{,}000$, where many substitute copies create substantial internal slack, to 65.6% at $N = 200$ under quadrupled demand, where baseline ranking drives the highest-demand type into exhaustion. Under base parameters, the gain is 42.7% at $N = 40$, 18.0% at $N = 100$, and 9.2% at $N = 200$.

In Regime A, the index sharply reduces exposure to products as they approach depletion. The baseline does not, and pays the difference in backorder cost. The gap narrows at $N = 1{,}000$ because many copies of each type create enough slack that even a static ranking can often find inventory. This mean-field concentration effect limits the large-N upside under the replicated-type design.

Regime B: Instantaneous Replenishment. Regime B reintroduces replenishment without allowing ranking and replenishment to overlap. Table 5 reports the same configurations as Table 4. The index retains a positive advantage in every configuration. The gains are smaller than in Regime A because instantaneous replenishment absorbs part of the downside of poor ranking: once a product reaches low inventory, the replenishment action immediately restores stock. The remaining gap therefore measures the value of allocating display positions more efficiently even when restocking is available.

TABLE 4. Regime A (no replenishment). Discounted cost per item J/N under the index and the common demand-weighted/margin-only baseline. Imp.% $= (J_{\text{bl}} - J_{\text{idx}})/J_{\text{bl}} \times 100$ is the cost reduction achieved by the index over the baseline.

	J_{idx}/N	J_{bl}/N	Imp.%†
Panel A: Catalog-size scaling (base parameters, 4 slots)			
$N = 40$	1.233	2.150	42.7
$N = 100$	1.725	2.103	18.0
$N = 200$	1.893	2.084	9.2
$N = 1{,}000$	2.028	2.067	1.9
Panel B: Stress scenarios ($N = 200$, 4 slots unless noted)			
High demand ($\theta \times 2$)	1.724	2.612	34.0
High demand ($\theta \times 4$)	1.389	4.038	65.6
Tight inventory ($K = 10$)	1.206	1.566	23.0
10 display slots	1.851	2.040	9.3
10 slots, $\theta \times 2$	1.641	2.574	36.2

† 100 Monte Carlo replications per cell. Standard errors on J/N are ≤ 0.027 at $N = 40$ and decay to ≤ 0.002 at $N = 1{,}000$. All reported improvements exceed twice the combined SE and are statistically significant at the 95% level.

The numerical levels in Regime B are lower than in Regime A because immediate replenishment limits the cost of stock depletion. The fact that the index advantage remains positive in every scenario is the main robustness point: even when the firm can restock immediately, there is value in ranking products according to inventory-adjusted exposure value rather than static demand or margin.

Finally, the computational cost is small. Across both regimes, each decision epoch requires only a lookup of product-level priority scores and a partial sort over N products.

6. Discussion and Conclusions

This paper studies how an online retailer should govern recommendation rankings when ranking affects both demand and operating costs. The focal managerial problem is economically important. Recommendation order does not merely reveal what consumers may want; it allocates attention across products. When products differ

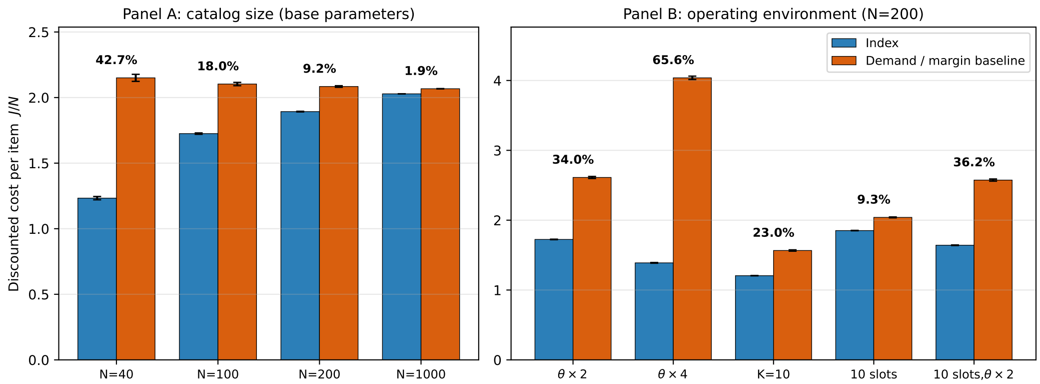

FIGURE 1. Regime A (no replenishment): discounted cost per item under the index lookup policy and the demand-weighted/margin-only baseline. Error bars are ± 1 Monte Carlo standard error over 100 replications. Numbers above each pair report the cost reduction $(J_{\text{bl}} - J_{\text{idx}})/J_{\text{bl}}$. Panel A (left) varies the catalog size N under base parameters; Panel B (right) fixes $N = 200$ and varies the operating environment.

TABLE 5. Regime B (instantaneous replenishment). Same scenarios and columns as Table 4. The reset action is retained but executed with zero hold time, so ranking and replenishment do not overlap.

	J_{idx}/N	J_{bl}/N	Imp.%[†]
Panel A: Catalog-size scaling (base parameters, 4 slots)			
$N = 40$	1.233	1.384	10.9
$N = 100$	1.725	1.790	3.6
$N = 200$	1.893	1.925	1.6
$N = 1{,}000$	2.028	2.035	0.3
Panel B: Stress scenarios ($N = 200$, 4 slots unless noted)			
High demand ($\theta \times 2$)	1.724	1.851	6.8
High demand ($\theta \times 4$)	1.389	1.755	20.8
Tight inventory ($K = 10$)	1.206	1.266	4.8
10 display slots	1.851	1.883	1.7
10 slots, $\theta \times 2$	1.641	1.770	7.2

[†] 100 Monte Carlo replications per cell. Standard errors are computed as in Table 4.

in margins, inventories, backlog costs, and replenishment costs, the firm should not rank only by relevance, demand potential, margin, or one-period expected revenue.

The relevant object is the inventory-adjusted shadow value of attention: the value generated by directing additional attention to a product net of the continuation-value cost of the inventory state it creates.

The technical problem is difficult because recommendation rankings generate a high-dimensional dynamic control problem. Products evolve stochastically as demand arrives, rankings couple products through the one-product-per-position constraint, and replenishment decisions interact with future ranking incentives. Exact dynamic programming is infeasible at realistic catalog scale, and related restless-bandit and queueing-control problems are known to be computationally hard (Papadimitriou and Tsitsiklis, 1999; Guha et al., 2010). We show that the recommendation problem can be formulated as a multi-mode restless-bandit problem and relaxed through a Whittle-style Lagrangian decomposition. The relaxation separates the system-level ranking problem into single-product subproblems indexed by a shadow price of discounted attention.

Our main analytical contribution is a closed-form attention index for each product and inventory state. The index prices the marginal value of assigning recommendation attention to a product at a given inventory level. It incorporates the demand created by attention, the margin earned from that demand, and the inventory-related costs that the induced demand creates. The ranking position itself enters through the attention weight α_k: the score from assigning product i to position k is $\alpha_k \psi_i(s_i)$. Thus the product-level index $\psi_i(s_i)$ measures the shadow value of attention for product i at inventory state s_i, while the assignment problem allocates higher-attention positions to products with higher attention values.

The theoretical structure has three parts. First, replenishment has a threshold form: below a product-specific inventory level, replenishment is optimal, so the reset region is downward closed. Second, the relaxed single-product problem admits a monotone partition over inventory states. Higher inventory states receive weakly more attention, while lower inventory states receive weakly less attention or are reset. Third, following the marginal-productivity-index methodology of Nino-Mora (2006);

Niño-Mora (2007), this partition structure yields a closed-form index for the relaxed problem. The resulting policy is analytically interpretable and computationally usable: the firm computes product-level attention values and solves a static assignment problem rather than an exponentially large dynamic program.

The computational experiments evaluate this mechanism in small systems, where exact dynamic programming provides a benchmark, and in large systems, where only scalable policies are implementable. The purpose of these experiments is not only to measure cost performance, but also to separate the two margins in the policy. The ranking margin asks whether products are assigned to positions according to their inventory-adjusted attention values. The replenishment margin asks whether a product should remain in the ranking system or be reset. This decomposition is useful managerially because ranking errors and replenishment errors have different operational interpretations.

The broader lesson is that recommendation systems are not only prediction systems. They are demand-shaping systems. This matters because demand-side and cost-side models are often analyzed separately. The economics and recommendation literatures emphasize attention, relevance, and demand response (Varian, 2007; Athey and Ellison, 2011; Linden et al., 2003; Koren et al., 2009); operations and revenue-management models emphasize inventory, capacity, substitution, and cost minimization (Gallego and Van Ryzin, 1994, 1997; Smith and Agrawal, 2000; Mahajan and Van Ryzin, 2001). In practice, however, the same managerial action often affects both sides. A recommendation can increase demand while also consuming scarce inventory, labor, service capacity, or production capacity. Our online retail model provides a tractable setting in which these forces can be studied jointly.

This logic extends beyond online retail. A restaurant server recommending a dish should account not only for customer fit and menu margin, but also for kitchen congestion and ingredient availability. A hotel recommending room upgrades should account for housekeeping constraints and future room availability. A sports club steering customers toward classes or premium services should account for instructor capacity, facility congestion, and already committed enrollments. In each case, the recommendation changes demand for an option whose cost depends on the current operating state. The managerial error is the same: treating recommendation as persuasion or prediction while ignoring the cost of serving the demand it creates.

The main implication is organizational. Firms should not separate the model that stimulates demand from the model that prices the cost of serving demand. Recommendation teams often optimize relevance, click-through, conversion, or short-run revenue, while operations teams absorb the consequences through stockouts, backlog, congestion, replenishment pressure, and service failures. Inventory-aware ranking provides one way to align these objectives. The ranking system should consume not only demand and margin information, but also inventory levels, backlog penalties, replenishment costs, and other operating-state variables. A product should be promoted when incremental attention has high inventory-adjusted value, not merely when predicted demand or immediate expected revenue is high.

In sum, recommendation order is a strategic control over scarce attention. When attention stimulates demand and demand changes future operating states, the firm must internalize the shadow value of the operating resources that demand uses. The index policy turns this economic logic into an implementable rule: allocate attention where it creates value net of future operating costs, reduce attention when demand stimulation would create costly scarcity, and replenish when resetting inventory dominates continued promotion.

Appendix A. Diminishing Discovery

Proof of Lemma 1. Adding one more inspected product cannot reduce the realized maximum, so $U(L)$ is increasing.

For concavity, define $Y_j = X_j^+$. Since

$$M_L = \max\{Y_1, \dots, Y_L\},$$

and since exchangeability is preserved by measurable transformations, the vector $(Y_1, \dots, Y_N)$ is exchangeable conditional on $\mathcal{H}$. Conditional on the unordered realized multiset $\{Y_1, \dots, Y_N\}$, exchangeability implies that the maximum of the first k inspected positions has the same distribution as the sample maximum from a sample of size k drawn without replacement from that finite population. This is the standard finite-population order-statistic representation of a sample maximum (see, e.g., David and Nagaraja, 2003; Arnold et al., 1992).

For any fixed finite population of realized residual values, the expected sample maximum is increasing and discretely concave in the sample size. An additional draw improves the current maximum only when it exceeds the best value already observed, and this improvement opportunity becomes weaker as the sample size grows. Equivalently, for each threshold z, the event that the sample maximum exceeds z is the event that the sample hits the upper contour set $\{j : Y_j > z\}$. The probability of hitting any fixed upper contour set under sampling without replacement has decreasing increments in the sample size; the expected maximum is obtained by integrating these threshold-crossing probabilities over z. Hence, conditional on the unordered realized multiset,

$$\begin{aligned} &\mathbb{E}[M_{k+1} - M_k \,|\, \mathcal{H}, \{Y_1, \dots, Y_N\}] \\ &\qquad \le \mathbb{E}[M_k - M_{k-1} \,|\, \mathcal{H}, \{Y_1, \dots, Y_N\}] \,. \end{aligned}$$

Taking expectations over the unordered multiset gives

$$U(k+1) - U(k) \le U(k) - U(k-1).$$

$\square$

Appendix B. Continuous-Time Optimal Control and Index Derivations

This appendix details the restless-bandit formulation and provides the proofs for the closed-form index used in Section 4. We focus on the single-arm analysis; the multi-product policy follows from the assignment decomposition.

B.1. Lagrangian Relaxation and Monotonicity.

Recall the priced single-arm Bellman equation for a given attention charge $\eta \geq 0$:

$$\begin{aligned} \beta v_\eta(s) = \min_{a\in\mathcal{A}} \Big\{ & C(s,a) + \eta r(a) \\ & + \sum_{s'} q(s' \mid s; a) v_\eta(s') \Big\}, \end{aligned} \tag{27}$$

where

$$\mathcal{A} = \{1, \dots, N, \mathcal{R}\}.$$

For ranking positions, the resource is recommendation attention,

$$r(a) = \alpha_a, \qquad a = 1, \dots, N,$$

and the translated reset control consumes no recommendation attention,

$$r(\mathcal{R}) = 0.$$

The optimal policy for this problem exhibits a monotone partition structure. First, if it is optimal to replenish at state s_j, then it is also optimal to replenish at s_{j-1}. Hence the reset region is downward closed, and we denote its largest state by $\bar{s}$.

Lemma 2 (Monotone partition structure)**.** *For every $\eta \geq 0$, there exists an optimal stationary deterministic policy. Moreover, there exists an optimal selector $A_\eta(s)$ such that, on the nonreset region, whenever $s_1 < s_2$,*

$$A_\eta(s_1), A_\eta(s_2) \in \{1, \dots, N\} \quad \Longrightarrow \quad A_\eta(s_1) \geq A_\eta(s_2).$$

Equivalently, the assigned attention level $\alpha_{A_\eta(s)}$ is weakly increasing in the inventory state s. Hence the state space decomposes into a replenishment threshold and contiguous ranking blocks on each of which the optimal ranking action is constant.

Proof. Existence of an optimal stationary deterministic policy is standard for the discounted finite-state CTMDP.

Fix $\eta \geq 0$, and write

$$\mu_a := \alpha_a \theta$$

for the demand rate under ranking position a. For a ranking action $a \in \{1, \ldots, N\}$, the Bellman action value at state s is

$$Q_\eta(s, a) := C(s, a) + \eta r(a) + \sum_{s'} q(s' \mid s; a) v_\eta(s').$$

Using

$$C(s, a) = c(s) - \mu_a m, \qquad r(a) = \alpha_a,$$

and

$$q(s - 1 \mid s; a) = \mu_a, \qquad q(s \mid s; a) = -\mu_a,$$

we obtain

$$Q_\eta(s, a) = c(s) + \alpha_a \eta - \alpha_a \theta m + \alpha_a \theta \{v_\eta(s - 1) - v_\eta(s)\}.$$

Define the backward difference

$$\Delta_\eta(s) := v_\eta(s) - v_\eta(s - 1).$$

Since the one-arm inventory cost $c(s) = hs_+ + p(-s)_+$ is convex, and the ranking dynamics move the state one step downward, standard value-iteration arguments imply that v_η is convex on the integer state space. Equivalently, $\Delta_\eta(s)$ is nondecreasing in s.

Now compare two ranking actions $1 \leq a < b \leq N$. Their action-value difference is

$$Q_\eta(s, a) - Q_\eta(s, b) = (\alpha_a - \alpha_b) \left[\eta - \theta\big(m + \Delta_\eta(s)\big)\right]. \tag{28}$$

Because $a < b$ implies $\alpha_a \geq \alpha_b$, and because $\Delta_\eta(s)$ is nondecreasing in s, the right-hand side of (28) is nonincreasing in s. Therefore each pair of ranking actions has the single-crossing property: once a lower-attention action is weakly preferred at some state, it remains weakly preferred at all lower states. Since ranking actions are totally ordered, the minimal optimal ranking selector is nonincreasing in s. Combining this with the replenishment threshold yields the stated partition structure. □

B.2. **Discounted Occupation Measures Under a Fixed Partition.** Fix a stationary partition policy S with maximal reset state $\bar{s}$. For each active state $y \in \{\bar{s}+1, \dots, K\}$, let $a_S(y)$ denote the ranking action used by S at y, and write

$$\lambda_y := \alpha_{a_S(y)}\theta.$$

Define the statewise flow cost under S by

$$c_S(y) = \begin{cases} c(y) - \lambda_y m, & y > \bar{s}, \\ T + b(K - \bar{s}), & y = \bar{s}, \end{cases}$$

and the statewise attention under S by

$$r_S(y) = \begin{cases} \alpha_{a_S(y)}, & y > \bar{s}, \\ 0, & y = \bar{s}. \end{cases}$$

For any initial state $x \in \{\bar{s}, \dots, K\}$, define the discounted occupation measure, for $y \in \{\bar{s}, \dots, K\}$, by

$$M_x^S(y) := \mathbb{E}_x^S\left[\int_0^\infty e^{-\beta t}\mathbf{1}\{X_t = y\}\, dt\right].$$

The fixed-policy discounted cost and attention are then

$$F_S(x) = \sum_{y=\bar{s}}^{K} c_S(y) M_x^S(y), \tag{29}$$

and

$$G_S(x) = \sum_{y=\bar{s}}^{K} r_S(y) M_x^S(y). \tag{30}$$

The discounted occupation measures satisfy the one-step recursions

$$M_x^S(y) = \frac{\mathbf{1}\{x = y\}}{\beta + \lambda_x} + \frac{\lambda_x}{\beta + \lambda_x} M_{x-1}^S(y), \qquad x > \bar{s}, \tag{31}$$

and

$$M_{\bar{s}}^S(y) = \frac{\mathbf{1}\{\bar{s} = y\}}{\beta + 1} + \frac{1}{\beta + 1} M_K^S(y). \tag{32}$$

Solving these recursions yields the explicit formulas below.

Proposition 2 (Closed form for the discounted occupation measures)**.** *For $y > \bar{s}$,*

$$M_K^S(y) = \frac{1}{\beta + \lambda_y} \prod_{t=y+1}^{K} \frac{\lambda_t}{\beta + \lambda_t} \times \left(1 - \frac{1}{\beta+1} \prod_{t=\bar{s}+1}^{K} \frac{\lambda_t}{\beta + \lambda_t}\right)^{-1}, \tag{33}$$

and

$$M_K^S(\bar{s}) = \frac{1}{\beta + 1} \prod_{t=\bar{s}+1}^{K} \frac{\lambda_t}{\beta + \lambda_t} \times \left(1 - \frac{1}{\beta+1} \prod_{t=\bar{s}+1}^{K} \frac{\lambda_t}{\beta + \lambda_t}\right)^{-1}. \tag{34}$$

More generally, for $x \in \{\bar{s}, \dots, K\}$ and $y > \bar{s}$,

$$M_x^S(y) = \frac{1}{\beta + \lambda_y} \prod_{t=y+1}^{x} \frac{\lambda_t}{\beta + \lambda_t} + \frac{1}{\beta+1} \prod_{t=\bar{s}+1}^{x} \frac{\lambda_t}{\beta + \lambda_t} M_K^S(y), \qquad \bar{s} < y \leq x. \tag{35}$$

If $y > x$, then

$$M_x^S(y) = \frac{1}{\beta+1} \prod_{t=\bar{s}+1}^{x} \frac{\lambda_t}{\beta + \lambda_t} M_K^S(y).$$

Moreover,

$$M_x^S(\bar{s}) = \frac{1}{\beta+1} \prod_{t=\bar{s}+1}^{x} \frac{\lambda_t}{\beta + \lambda_t} \times \big(1 + M_K^S(\bar{s})\big). \tag{36}$$

Proof. Equation (33) is obtained by setting $x = K$ in (31) and closing the cycle through (32). Equation (34) follows similarly. For general x, repeated substitution of (31) gives the first-pass contribution down to y, and (32) appends the regenerative continuation from K, yielding (35), the displayed formula for $y > x$, and (36). $\square$

As a consequence, the recursions for F_S and G_S are obtained directly from (29)–(30) and (31):

$$F_S(x) = \frac{c_S(x) + \lambda_x F_S(x-1)}{\beta + \lambda_x}, \qquad x > \bar{s}, \tag{37}$$

$$G_S(x) = \frac{r_S(x) + \lambda_x G_S(x-1)}{\beta + \lambda_x}, \qquad x > \bar{s}, \tag{38}$$

with reset boundary

$$\begin{aligned} F_S(\bar{s}) &= \frac{T + b(K-\bar{s}) + F_S(K)}{\beta+1}, \\ G_S(\bar{s}) &= \frac{G_S(K)}{\beta+1}. \end{aligned} \tag{39}$$

B.3. **One-Shot Deviation Index.** Fix a state $s > \bar{s}$, and let the background action under S at s be

$$a := a_S(s), \qquad \mu_a := \alpha_a \theta.$$

For an alternative ranking action $k \in \{1, \dots, N\}$, define

$$\mu_k := \alpha_k \theta.$$

The first-exit identities are

$$F_{(k,S)}(s) = \frac{C(s,k) + \mu_k F_S(s-1)}{\beta + \mu_k}, \tag{40}$$

$$G_{(k,S)}(s) = \frac{\alpha_k + \mu_k G_S(s-1)}{\beta + \mu_k}, \tag{41}$$

and, for reset,

$$F_{(\mathcal{R},S)}(s) = \frac{T + b(K-s) + F_S(K)}{\beta+1}, \tag{42}$$

$$G_{(\mathcal{R},S)}(s) = \frac{G_S(K)}{\beta+1}. \tag{43}$$

Lemma 3 (Rank-to-rank marginal attention)**.** *Fix a continuation policy S and an active state $s > \bar{s}$. Suppose S assigns position a at s. Then, for every ranking position*

k,

$$G_S(s) - G_{(k,S)}(s) = \frac{\beta(\alpha_a - \alpha_k)}{\beta + \alpha_a\theta} \times \frac{1 + \theta G_S(s-1)}{\beta + \alpha_k\theta}.$$

Consequently, $G_S(s) - G_{(k,S)}(s) > 0$ iff $\alpha_a > \alpha_k$, equals zero iff $\alpha_a = \alpha_k$, and is negative iff $\alpha_a < \alpha_k$.

Proof. Write

$$g = G_S(s-1).$$

Under the attention-resource normalization $r(j) = \alpha_j$,

$$G_S(s) = \frac{\alpha_a + \alpha_a\theta g}{\beta + \alpha_a\theta}.$$

The one-shot deviation to position k gives

$$G_{(k,S)}(s) = \frac{\alpha_k + \alpha_k\theta g}{\beta + \alpha_k\theta}.$$

Subtracting,

$$\begin{aligned} G_S(s) - G_{(k,S)}(s) &= \frac{\alpha_a + \alpha_a\theta g}{\beta + \alpha_a\theta} - \frac{\alpha_k + \alpha_k\theta g}{\beta + \alpha_k\theta} \\ &= (1 + \theta g)\left[\frac{\alpha_a}{\beta + \alpha_a\theta} - \frac{\alpha_k}{\beta + \alpha_k\theta}\right] \\ &= \frac{\beta(\alpha_a - \alpha_k)}{\beta + \alpha_a\theta} \\ &\quad \times \frac{1 + \theta g}{\beta + \alpha_k\theta}. \end{aligned}$$

All factors other than $\alpha_a - \alpha_k$ are positive. The sign conclusion follows, the negativity cancels fully in the complete index as we show below. □

Proposition 3 (Ranking-to-ranking critical charge)**.** *For every $k \in \{1, \ldots, N\}$ with $\alpha_k \neq \alpha_a$,*

$$\nu(k, s; S) = \psi(s; S),$$

where

$$\psi(s; S) = \frac{\theta\{c(s) + \beta m - \beta F_S(s-1)\}}{\beta\{1 + \theta G_S(s-1)\}}. \tag{44}$$

Equivalently, using the occupation-measure representation,

$$\psi(s;S) = \theta\left[c(s) + \beta m - \beta \sum_{y=\bar{s}}^{K} c_S(y) M_{s-1}^{S}(y)\right] \times \left[\beta + \beta\theta \sum_{y=\bar{s}}^{K} r_S(y) M_{s-1}^{S}(y)\right]^{-1}. \tag{45}$$

If $\alpha_k = \alpha_a$, the two positions are attention-equivalent for this comparison.

Proof. Using (40),

$$F_{(k,S)}(s) - F_S(s) = \frac{C(s,k) + \mu_k F_S(s-1)}{\beta + \mu_k} - \frac{C(s,a) + \mu_a F_S(s-1)}{\beta + \mu_a}.$$

Since

$$C(s,k) - C(s,a) = (\alpha_a - \alpha_k)\theta m,$$

direct simplification gives

$$F_{(k,S)}(s) - F_S(s) = (\alpha_a - \alpha_k)\theta \times \{c(s) + \beta m - \beta F_S(s-1)\} \times \left[(\beta + \alpha_k\theta)(\beta + \alpha_a\theta)\right]^{-1}.$$

Similarly, using (41),

$$G_S(s) - G_{(k,S)}(s) = \beta(\alpha_a - \alpha_k)\{1 + \theta G_S(s-1)\} \times \left[(\beta + \alpha_k\theta)(\beta + \alpha_a\theta)\right]^{-1}.$$

When $\alpha_k \neq \alpha_a$, both marginal differences contain the same rank-comparison factor

$$(\alpha_a - \alpha_k)\left[(\beta + \alpha_k\theta)(\beta + \alpha_a\theta)\right]^{-1}.$$

Substituting the two displayed expressions this factor cancels, so

$$\nu(k,s;S) = \frac{\theta\{c(s) + \beta m - \beta F_S(s-1)\}}{\beta\{1 + \theta G_S(s-1)\}} = \psi(s;S).$$

Substituting (29)–(30) gives (45). □

Reset consumes no current attention, but its continuation value is evaluated under the post-reset state K.

Proposition 4 (Ranking-versus-reset critical charge in occupation-measure form)**.** *Suppose the background action at $s > \bar{s}$ is ranking. Then, whenever the denominator is nonzero,*

$$\begin{aligned}\nu(\mathcal{R}, s; S) = \Bigg[& T + b(K-s) + \sum_{y=\bar{s}}^{K} c_S(y) M_K^S(y) \\ & - (\beta+1) \sum_{y=\bar{s}}^{K} c_S(y) M_s^S(y) \Bigg] \\ \times \Bigg[& (\beta+1) \sum_{y=\bar{s}}^{K} r_S(y) M_s^S(y) \\ & - \sum_{y=\bar{s}}^{K} r_S(y) M_K^S(y) \Bigg]^{-1}. \end{aligned} \tag{46}$$

The reset boundary is obtained from the corresponding charged indifference equation.

Proof. Substitute (29)–(30) into (42)–(43) and then into (**??**). □

B.4. **Blockwise Switching Equations.** Let

$$B_j = \{\ell_j, \dots, u_j\}$$

be a ranking block of S on which the action is constantly a_j, and let

$$s = \ell_j + n - 1 \in B_j, \qquad n = 1, \dots, u_j - \ell_j + 1.$$

Define

$$q_j := \frac{\alpha_{a_j}\theta}{\beta + \alpha_{a_j}\theta}.$$

On a sign-homogeneous block, the terms entering the attention index have simple geometric recursions.

Proposition 5 (Block recursions for the index terms)**.** *If $B_j \subseteq \{1,\dots,K\}$, so $c(t) = ht$ throughout the block, then*

$$hs + \beta m - \beta F_S(s-1) = q_j^{n-1}\big[h\ell_j + \beta m - \beta F_S(\ell_j - 1)\big] + \frac{h+\beta m}{1-q_j}\big(1 - q_j^{n-1}\big),$$

and

$$\beta\{1 + \theta G_S(s-1)\} = q_j^{n-1}\beta\{1+\theta G_S(\ell_j - 1)\} + \frac{\beta}{1-q_j}\big(1-q_j^{n-1}\big).$$

If $B_j \subseteq \{\bar{s}+1,\dots,0\}$, so $c(t) = -pt$ throughout the block, then

$$-ps + \beta m - \beta F_S(s-1) = q_j^{n-1}\big[-p\ell_j + \beta m - \beta F_S(\ell_j - 1)\big] + \frac{-p+\beta m}{1-q_j}\big(1 - q_j^{n-1}\big),$$

and

$$\beta\{1 + \theta G_S(s-1)\} = q_j^{n-1}\beta\{1+\theta G_S(\ell_j - 1)\} + \frac{\beta}{1-q_j}\big(1-q_j^{n-1}\big).$$

Proof. On block B_j, the fixed-policy recursions are

$$F_S(t) = q_j F_S(t-1) + \frac{c(t) - \alpha_{a_j}\theta m}{\beta + \alpha_{a_j}\theta},$$

and

$$G_S(t) = q_j G_S(t-1) + \frac{\alpha_{a_j}}{\beta + \alpha_{a_j}\theta}.$$

For a positive block, define the current index numerator term at state $t+1$ by

$$h(t+1) + \beta m - \beta F_S(t).$$

The recursion above implies that this term equals

$$h + \beta m + q_j\{ht + \beta m - \beta F_S(t-1)\}.$$

Iterating from $t = \ell_j$ to $t = s-1$ gives the first displayed formula. The negative-block formula follows identically with h replaced by $-p$.

For the denominator term, the attention recursion implies

$$1 + \theta G_S(t) = 1 + q_j\{1 + \theta G_S(t-1)\}.$$

Multiplying by β and iterating gives the displayed denominator formula in both cases.

□

The ranking switching equation on an active state is

$$\eta = \psi(s; S).$$

Using (44), this is equivalent to

$$\eta\,\beta\{1 + \theta G_S(s-1)\} = \theta\{c(s) + \beta m - \beta F_S(s-1)\}.$$

On a positive block, this becomes

$$\eta\,\beta\{1 + \theta G_S(s-1)\} = \theta\{hs + \beta m - \beta F_S(s-1)\}.$$

On a negative block, it becomes

$$\eta\,\beta\{1 + \theta G_S(s-1)\} = \theta\{-ps + \beta m - \beta F_S(s-1)\}.$$

By Proposition 5, each side is affine in q_j^{n-1} after substitution.

The reset boundary is determined by the charged indifference equation

$$F_S(s) + \eta G_S(s) = \frac{T + b(K-s) + F_S(K) + \eta G_S(K)}{\beta + 1}. \tag{47}$$

In particular, the first active state $\bar{s}+1$ is consistent with charged optimality precisely when ranking at $\bar{s}+1$ weakly dominates reset under (47).

Lemma 4 (Uniqueness of the ranking threshold)**.** *Fix a sign-homogeneous block B_j. Then the switching equation*

$$\eta = \psi(s; S)$$

has at most one solution $s \in B_j$, except in the degenerate case in which the equation is identically satisfied on the whole block.

Proof. Let $s = \ell_j + n - 1$. By Proposition 5, both

$$c(s) + \beta m - \beta F_S(s-1)$$

and

$$\beta\{1 + \theta G_S(s-1)\}$$

are affine functions of q_j^{n-1} on a sign-homogeneous block. Therefore the switching equation

$$\eta\, \beta\{1 + \theta G_S(s-1)\} = \theta\{c(s) + \beta m - \beta F_S(s-1)\}$$

is affine in q_j^{n-1}. Since $0 < q_j < 1$, the map $n \mapsto q_j^{n-1}$ is strictly monotone. Hence the equation has at most one solution unless the affine expression is identically zero on the block. $\square$

Accordingly, the single-arm partition is recovered by solving the reset-entry condition (47) together with the ranking switching equation $\eta = \psi(s; S)$ block by block.

B.5. **Optimal Assignment.** We now combine the single-arm objects into the multi-product policy induced by the relaxation. For each product i, let S_i^* denote the single-product continuation policy used to compute its attention index, and let $\bar{s}_i$ be its maximal reset state. At global state

$$\boldsymbol{s} = (s_1, \dots, s_N),$$

define the reset set and active set by

$$\mathcal{J}(\boldsymbol{s}) := \{i : \; s_i \le \bar{s}_i\}, \qquad \mathcal{I}(\boldsymbol{s}) := \{i : \; s_i > \bar{s}_i\}.$$

Products in $\mathcal{J}(\boldsymbol{s})$ take the reset action. Products in $\mathcal{I}(\boldsymbol{s})$ participate in the ranking assignment.

For an active product $i \in \mathcal{I}(\boldsymbol{s})$, define

$$\psi_i(s_i) = \psi_i(s_i; S_i^*),$$

where the right-hand side is computed from (44) using product i's primitives and partition policy S_i^*. The score from assigning product i to position k is

$$\alpha_k \psi_i(s_i).$$

Thus $\psi_i(s_i)$ is the product-state shadow value of attention, while α_k is the amount of attention carried by position k.

Theorem 1 (Attention assignment)**.** *At any global state $\boldsymbol{s}$, the attention-index policy assigns ranking positions to active products by solving*

$$
\begin{aligned}
(48) \qquad \max_{x_{ik}\in\{0,1\}} \quad & \sum_{i\in\mathcal{I}(\boldsymbol{s})} \psi_i(s_i) \sum_{k=1}^{N} \alpha_k x_{ik} \\
\text{s.t.} \quad & \sum_{k=1}^{N} x_{ik} = 1, \qquad i \in \mathcal{I}(\boldsymbol{s}), \\
& \sum_{i\in\mathcal{I}(\boldsymbol{s})} x_{ik} \le 1, \qquad k = 1, \dots, N.
\end{aligned}
$$

Equivalently, each active product receives one ranking position, and each ranking position is assigned to at most one active product. When every product is active, this reduces to the equality-constraint assignment formulation used in Section 4.

Proof. The Lagrangian relaxation decouples the dynamic optimization across products. For each active product i, the single-product index $\psi_i(s_i)$ is the critical price of discounted attention at the current inventory state. Assigning that product to position k supplies attention α_k, so the product–position contribution is $\alpha_k \psi_i(s_i)$. The remaining system-level constraint is the one-copy-per-position constraint. Therefore the feasible index implementation is exactly the assignment problem (48). Products in $\mathcal{J}(\boldsymbol{s})$ take reset, while products in $\mathcal{I}(\boldsymbol{s})$ are ranked. □

Since $\alpha_1 \ge \cdots \ge \alpha_N$, the assignment has the usual sorting interpretation when all active products must be ranked: higher-attention positions are assigned to products with larger attention indices. The explicit assignment formulation is retained because it also handles ties, reset states, and variants with fewer visible positions.

References

Adomavicius, G. and A. Tuzhilin (2005): "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE transactions on knowledge and data engineering*, 17, 734–749.

Aouad, A., V. Farias, and R. Levi (2021): "Assortment optimization under consider-then-choose choice models," *Management Science*, 67, 3368–3386.

Aouad, A., R. Levi, and D. Segev (2018): "Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences," *Operations Research*, 66, 1321–1345.

——— (2019): "Approximation algorithms for dynamic assortment optimization models," *Mathematics of Operations Research*, 44, 487–511.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992): *A First Course in Order Statistics*, New York: John Wiley & Sons.

Athey, S. and G. Ellison (2011): "Position auctions with consumer search," *The Quarterly Journal of Economics*, 126, 1213–1270.

Bai, Y., O. El Housni, P. Rusmevichientong, and H. Topaloglu (2025): "Coordinated Inventory Stocking and Assortment Customization," *Operations Research*.

Bernstein, F., A. G. Kök, and L. Xie (2015): "Dynamic assortment customization with limited inventories," *Manufacturing & Service Operations Management*, 17, 538–553.

Bimpikis, K., O. Candogan, and D. Saban (2019): "Spatial pricing in ride-sharing networks," *Operations Research*, 67, 744–769.

Bitran, G. R. and J.-C. Ferrer (2007): "On pricing and composition of bundles," *Production and Operations Management*, 16, 93–108.

Cachon, G. P., K. M. Daniels, and R. Lobel (2017): "The role of surge pricing on a service platform with self-scheduling capacity," *Manufacturing & Service Operations Management*, 19, 368–384.

Caro, F. and J. Gallien (2007): "Dynamic assortment with demand learning for seasonal consumer goods," *Management science*, 53, 276–292.

Chen, X., W. Ma, D. Simchi-Levi, and L. Xin (2024): "Assortment planning for recommendations at checkout under inventory constraints," *Mathematics of Operations Research*, 49, 297–325.

David, H. A. and H. N. Nagaraja (2003): *Order Statistics*, Hoboken, NJ: Wiley, 3 ed.

De Biasio, A., D. Jannach, and N. Navarin (2024): "Model-Based Approaches to Profit-Aware Recommendation," *Expert Systems with Applications*, 249, 123642.

De Biasio, A., A. Montagna, F. Aiolli, and N. Navarin (2023): "A Systematic Review of Value-Aware Recommender Systems," *Expert Systems with Applications*, 226, 120131.

Demirezen, E. M. and S. Kumar (2016): "Optimization of recommender systems based on inventory," *Production and Operations Management*, 25, 593–608.

Dong, L., P. Kouvelis, and Z. Tian (2009): "Dynamic pricing and inventory control of substitute products," *Manufacturing & Service Operations Management*, 11, 317–339.

Erginbas, Y. E., S. Phade, and K. Ramchandran (2023): "Interactive Learning with Pricing for Optimal and Stable Allocations in Markets," in *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, 9773–9806.

Feltham, G. A. and J. Xie (1994): "Performance measure congruity and diversity in multi-task principal/agent relations," *Accounting review*, 429–453.

Gallego, G. and G. Van Ryzin (1994): "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management science*, 40, 999–1020.

——— (1997): "A multiproduct dynamic pricing problem and its applications to network yield management," *Operations research*, 45, 24–41.

Glazebrook, K., D. Ruiz-Hernandez, and C. Kirkbride (2006): "Some indexable families of restless bandit problems," *Advances in Applied Probability*, 643–672.

Golrezaei, N., H. Nazerzadeh, and P. Rusmevichientong (2014): "Real-time optimization of personalized assortments," *Management Science*, 60, 1532–1551.

Guha, S., K. Munagala, and P. Shi (2010): "Approximation Algorithms for Restless Bandit Problems," *Journal of the ACM (JACM)*, 58, 3.

Holmstrom, B. and P. Milgrom (1991): "Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design," *The Journal of Law, Economics, and Organization*, 7, 24–52.

Honhon, D., V. Gaur, and S. Seshadri (2010): "Assortment planning and inventory decisions under stockout-based substitution," *Operations research*, 58, 1364–1379.

Koren, Y., R. Bell, and C. Volinsky (2009): "Matrix factorization techniques for recommender systems," *Computer*, 42, 30–37.

Lahaie, S. (2008): "An Analysis of Alternative Slot Auction Designs for Sponsored Search," in *Proceedings of the 7th ACM Conference on Electronic Commerce*.

Li, J., Y. Luo, and X. Zhang (2024): "Seesaw Experimentation: A/B Tests with Spillovers," *arXiv preprint arXiv:2411.02085*.

Li, J., G. Pisano, Y. Xu, and F. Zhu (2023): "Marketplace scalability and strategic use of platform investment," *Management science*, 69, 3958–3975.

Linden, G., B. Smith, and J. York (2003): "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, 7, 76–80.

Mahajan, S. and G. Van Ryzin (2001): "Inventory competition under dynamic consumer choice," *Operations Research*, 49, 646–657.

Miao, S. and X. Chao (2021): "Dynamic joint assortment and pricing optimization with demand learning," *Manufacturing & Service Operations Management*, 23, 525–545.

Musto, C., P. Lops, M. de Gemmis, and G. Semeraro (2021): "Context-aware graph-based recommendations exploiting personalized pagerank," *Knowledge-Based Systems*, 216, 106806.

Niño-Mora, J. (2001): "Restless Bandits, Partial Conservation Laws and Indexability," *Advances in Applied Probability*, 33, 76–98.

Nino-Mora, J. (2006): "Restless bandit marginal productivity indices, diminishing returns, and optimal control of make-to-order/make-to-stock M/G/1 queues," *Mathematics of Operations Research*, 31, 50–84.

Niño-Mora, J. (2007): "Dynamic priority allocation via restless bandit marginal productivity indices," *Top*, 15, 161–198.

Papadimitriou, C. H. and J. N. Tsitsiklis (1999): “The Complexity of Optimal Queuing Network Control,” *Mathematics of Operations Research*, 24, 293–305.

Piunovskiy, A. and Y. Zhang (2020): “Continuous-time Markov decision processes,” *Probability Theory and Stochastic Modelling*.

Rusmevichientong, P. and H. Topaloglu (2012): “Robust assortment optimization in revenue management under the multinomial logit choice model,” *Operations research*, 60, 865–882.

Ryzin, G. v. and S. Mahajan (1999): “On the relationship between inventory costs and variety benefits in retail assortments,” *Management Science*, 45, 1496–1509.

Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen (2007): “Collaborative filtering recommender systems,” in *The adaptive web: methods and strategies of web personalization*, Springer, 291–324.

Smith, S. A. and N. Agrawal (2000): “Management of multi-item retail inventory systems with demand substitution,” *Operations Research*, 48, 50–64.

Talluri, K. T. and G. J. Van Ryzin (2006): *The theory and practice of revenue management*, vol. 68, Springer Science & Business Media.

Topaloglu, H. (2013): “Joint stocking and product offer decisions under the multinomial logit model,” *Production and Operations Management*, 22, 1182–1199.

Varian, H. R. (2007): “Position auctions,” *international Journal of industrial Organization*, 25, 1163–1178.

Verloop, I. M. (2016): “Asymptotically Optimal Priority Policies for Indexable and Nonindexable Restless Bandits,” *Annals of Applied Probability*, 26, 1947–1995.

Weber, R. R. and G. Weiss (1990): “On an index policy for restless bandits,” *Journal of applied probability*, 27, 637–648.

Whittle, P. (1988): “Restless bandits: Activity allocation in a changing world,” *Journal of applied probability*, 287–298.